Furiously Fast
On the speed of change in formulaic language

Andreas Buerki
Centre for Language and Communication Research, Cardiff University, Wales

Addressing a topic that has been marginal to discussions within historical linguistics, this study looks at how extent and speed of language change can be quantified meaningfully using corpus data. Looking specifically at formulaic language (understood here as word sequences that instantiate typical phrasings), a solidly data-based assessment of the speed of change within a 100-year time window is offered. This includes both a relative determination of speed (against the speed of change in lexis which is generally thought to be the fastest type of linguistic change, cf. Algeo 1980:264; Trask and Millar 2010:7) as well as a new independent measure of speed which is easy to interpret and therefore of high validity, while also robust and potentially applicable to any linguistic feature that can be counted in corpus data. Using data from a diachronic reference corpus of 20th century German, it is shown that change in formulaic language is very notably faster than lexical change, that the extent of change over a century is comparable in extent to contemporary inter-genre variation and that overall, the rate of change does fluctuate somewhat at the level of temporal granularity employed in this study. It is also argued that quantifying the speed of linguistic change can play an important role in building a deeper understanding language change in general.

Keywords: historical linguistics, language change, speed of change, formulaic language, multi-word units
Word count: 9,500

1 Introduction

The speed or slowness of language change is an intriguing topic that has received marginal treatment in historical linguistics. It is intriguing because the rapidity of language change is at the same time a concept that resonates with ordinary language users to a degree where it seems that it would be one of the first questions historical linguists should answer but it is also a task that is highly complex – to the extent that it has so far largely eluded definitive treatment. Where assessments have been made, they have been variously based on fairly vague impressions (as in Fodor 1965:16; Bynon 1977:2, Mair 2006:34) or regarding a small set of choice examples studied in detail (changes in small sets of core vocabulary, e.g. Swadesh, 1955) or in very particular settings (such as work on sound change within particular communities, e.g. Labov, 1972). There are of course reasons why so far comparatively little research has gone into quantifying the speed of language change more generally and robustly, based on corpus linguistic evidence: it is unclear how speed should be measured, what it could be compared to, whether it is at all meaningful to measure an overall speed, or even speeds for different sub-systems of the language, whether speed is fairly constant or highly variable across different time periods and circumstances or whether indeed speed of change can only ever be meaningful on an item by item, setting by setting, time period by time period basis.

Approaching this topic via the speed of change in formulaic language (FL) holds at least two key advantages: despite a real surge in interest in FL and related

phenomena, change in FL has to date not benefitted from nearly as much diachronic research as syntax, morphology, lexis or the sound system. Given the increasingly recognised importance of FL in the functioning of language, this state of affairs is unfortunate and attaining an enhanced understanding of the diachronic behaviour of FL in an empirically well-founded manner is an important step toward rectifying it. Extent and speed of change reveal important aspects of the nature of FL, but a further key advantage is that looking at FL turns out to challenge long-held understandings regarding far more general aspects of linguistic change and therefore adds important insight to the study of language change in general.

In this investigation, the view on change is a bird's eye view: in pursuing questions after the extent and speed of change, the focus is on an overall quantification of change in the FL of a speech community over the 20th century rather than the analysis of particular items of FL and any individual changes they might undergo. The focus on broad quantification is not because particulars and exemplification are not important (they are), but because at this time in historical linguistics 'the big problem [...] is looking for generalities and reproducible results' (Bauer 1994:32, in relation to lexical change) rather than collecting hand-picked examples from which to extrapolate.

In the following, I first set out the understanding of FL in the context of this study. Then, after an overview of existing work on diachronic change and synchronic variation in FL and related work on quantifying language change more broadly, the method of investigation is laid out. This is followed by a presentation of results. Implications are considered in a final section.

2 Background

2.1 Formulaic Language

Common turns of phrase, variously labelled FL, phraseology, multi-word expressions, prefabs or similar, are expressions, typically longer than single words, that represent common ways of putting things in a speech community (cf. Burger et al., 1982: 1; Coulmas, 1979; Erman and Warren, 2000; Fillmore et al., 1988; Howarth, 1998: 25; Langacker, 2008: 84; Pawley, 2001). They may include such phenomena as collocations (e.g. *hugely disappointing*, *hardly surprising*, *strong coffee, file for a divorce* etc.), multi-word units (e.g. *open letter*, *the single market, cease and desist,* etc.), formulae proper (e.g. *in other words*, a formula introducing a paraphrase, or *in summary* to introduce conclusions, or *thank you for holding, your call is important to us*, a formula signalling that a call is in a queue), idioms (e.g. *live to tell the tale*), and very many other usual sequences (e.g. *just about, nothing short of X, in the [late] Xth Century,* etc.). While most treatments see these word sequences as predominantly lexically fixed constructions, albeit with variable elements (as indicated by the Xs in above examples), there are different views on how exactly FL should be delimited and described. Traditional phraseology, as well as treatments in Natural Language Processing, have tended to place particular emphasis on expressions that show sematic or formal irregularity (e.g. *red tape* is not tape that is red), though recently, this has become less rigid (Burger et al. 2007:11). More psycholinguistic treatments have taken the manner of mental processing as key (e.g. ‘a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use’ Wray, 2002:9; similarly Sinclair, 1991:110). A third strand of research

that could be labelled corpus linguistic and is followed in the present work, has tended to see the essence of FL in its recurrent usage in language produced by a speech community (e.g. Buerki 2016:18; Bybee, 2010: 35). Despite these differences in emphasis, there is broad agreement on the importance of the phenomenon of FL within language, both in terms of the proportion of a text that consists of FL (where estimates vary between 16% (Van Lancker-Sidtis and Rallon 2004) and over 80% (Altenberg 1998), depending on text type and definition) and the role of FL in accounting for processes in first language acquisition (Dąbrowska, 2014; Lieven and Brandt, 2011), L2 learning (Allerton, 1984; Bally, 1909: 70–3; Boers et al., 2006; Jespersen, 1904; Myles, 2004; Nattinger and DeCarrico, 1992; Pawley and Syder, 1983; Sorhus, 1977; Wray and Perkins, 2000), facilitating successful communication (Erman, 2007: 26; Feilke, 1994; 2003: 213; Wray, 2008: 20–1) and fluency (e.g. Wray, 2002: 35–7), beside many more areas.

2.2 Change in FL

Despite some interesting recent contributions to the study of change in FL (cf. Hyland and Jiang, 2018; Kopaczyk, 2012), research on quantifying change in FL in general is to date very scarce. Most existing research has focussed on individual items of FL or fairly small groups thereof. Nevertheless, some studies have commented on extent and speed of change in FL: Pei, for example, found that 'many of our word-combinations are of *recent* military origin' (1953:118 as quoted in Bauer 1994:29, my emphasis), citing examples such as *scorched earth, lend-lease, walkie talkie* or the univerbated *blackout, dogfight* and *blockbuster*. Similarly, Bischof, in a corpus study of collocations of emotion in Old, Middle and Modern French finds that '[m]any of the analysed Modern French collocations appear relatively late in the language, even in the 20th century' (2008:17). Burger and Linke (1998:750) conclude that '[w]ord sequences that have become phraseologically connected continue to develop in the course of language history [...]. They do so much like individual words, *but probably at higher speed.* (Burger and Linke, 1998:750, my emphasis, my translation). Hyland and Jiang (2018) mention that results of their study of lexical bundles across 50 years of journal articles in four fields, 'show a significant shift in uses' (2018:402). Discourse analytical work on FL has further shown that items of FL can undergo perceptible shifts in usage over relatively short periods (e.g. Bubenhofer, 2009; Stubbs, 2002). These comments suggest interesting goings-on, including that overall change in FL may be fairly rapid. To date, however, this notion has not yet been suitably substantiated, discussed or supported with the kind of large-scale data analysis that would appear to be necessary and so, as Bauer remarks in relation to Pei (1953), more questions are raised than answered: 'How many word combinations are involved [in recent change] […]? In what way are the examples typical?' (Bauer 1994:29).

2.3 Synchronic Variation in FL

By contrast to diachronic change, research on synchonic variation in FL is well-developed and has shown very clearly that items of FL vary widely across genres and registers (recent studies include Ädel and Erman, 2012:81; Biber, 2006; 2009; Biber and Barbieri, 2007; Biber, Conrad, and Cortes, 2003; Kuiper, 2009 but see also Biber et al., 1999; Burger and Buhofer, 1981). In the synchronic context, variation has been quantified in a general manner (i.e., beyond the study of individual items of FL or

small groups thereof). Gries (2010), for example, used clustered average G-gravity scores of bigrams in 19 sub-registers of the BNC-Baby corpus to accurately re-construct the four main register divisions in the corpus. This indicates once more that FL and similar phenomena are indicators of register. It also shows that general quantification is not only possible but can be meaningful (in this case for assigning texts to registers). In another example of research employing general quantification of synchronic FL-variation, Biber, Conrad, and Cortes (2004) compared lexical bundles in conversation, university classroom teaching, textbooks and academic prose. Three comparisons were made: FL-density (types and tokens), proportions of structural types (NP/PP-based, dependent clauses, VP-based) and functional types (referential, discourse organizers, stance) across the four registers. These not only provide comprehensive and therefore reliable characterisations but also very helpfully inform later, more detailed analyses of individual high-frequency items of FL. Thus general quantification has been a useful way of looking at synchronic variation in FL.

2.4 Quantifying language change

A third area to consider are quantifications of linguistic change in areas outside of FL. Early attempts in lexicostatistics and glottochronology (Swadesh, 1955; 1959; Sankoff, 1970) sought to take change in basic vocabulary (which was thought to operate at a stable rate of approximately 20% of words being replaced in 1,000 years, cf. Crowley and Bowern, 2010:148) to establish the time when two languages diverged. These methods enjoyed limited acceptance for a time but are no longer thought to be reliable due to their precarious methodological assumptions (cf. Bynon, 1977: 266-72; Crowley and Bowern 2010: 149-51 for discussion). Subsequent attempts to measure the speed of change of various linguistic phenomena, such as the rate of change index suggested by Johnson (1976) have not enjoyed broad acceptance, although lexicostatistic methods live on in strands of work on linguistic and cultural evolution concerned with the question of whether the rate of change in language is regular and independent of environment and population or indeed dependent on variables like population size. Some continue to employ rates of change in basic vocabulary lists (so-called *Swadesh lists*) to extrapolate to the rate of language change at large (Bromham et al., 2015; Wichmann and Holman, 2009), while other work uses phoneme inventory size (Moran et al, 2012) or various feature lists and databases of selected features (e.g. Greenhill et al., 2018, Nettle 1999). Using changes in limited item lists to model overall rates of change inevitably stretches validity, a circumstance acknowledged by proponents who insist (curiously to the linguist) that although 'models used should have a certain degree of realism, [they] should not try to imitate a complicated reality' (Wichmann, 2008:445). Consequently, how individual items should be weighed, how a highly specific sample of change can represent items and levels of analysis not studied are some of the issues left unexplored. An interesting direction is indicated by Sandøy (2009). In a study that relates the recent history of migration of various Norwegian communities to the rate of morphological change in their dialects, 17 morphological changes over 80 years (collected by various sociolinguistic studies) are weighed by their frequency in language use to arrive at coefficients of morphological change that can be compared. But even here, as Sandøy points out, not only can theoretical views impinge on what constitutes a change for authors of different studies (2009:286), but it is difficult to assess how accidental or complete a compiled list of (in this case morphological) changes is (2009:291), with unknown knock-on effects on results.

There are also studies that have shown it is possible to meaningfully and reliably quantify linguistic change, even over short periods of time (cf. Altintas, Can, and Patton, 2007; Chesley and Baayen, 2010; Juola, 2003; 2005). These studies have applied to diachronic data computational and mathematical methods, many of which were originally developed for authorship identification or stylometry (cf. e.g. Juola, 2007; Juola and Baayen, 2005). Differing research interests and traditions in this area have meant that some of the results appear, from a linguistic point of view, underanalysed or without clear linguistic relevance, while others offer new insight of potentially very significant import to diachronic studies of language.

In Altintas et al. (2007), for example, translations into Turkish of a number of classical works were analysed. For each of the works, two translations into Turkish were available. The earlier translations (from the mid-twentieth century) were compared to the later translations (from the end of the twentieth century) and morphological differences analysed quantitatively. This showed an increase in morphological complexity coupled with a 'decrease in stem level vocabulary richness' (Altintas et al. 2007:386). This observation was correlated with an independently documented tendency, over the period of observation, to replace foreign word stems with native Turkish creations. 'Such neologisms are usually obtained by adding suffixes to Turkish stems' (Altintas et al. 2007:386). The study concluded that 'in contemporary Turkish, one would use more suffixes to compensate for the fewer stems to preserve the expressive power of the language' (Altintas et al. 2007:386). In this manner, Altintas et al. empirically demonstrate a broad change across Turkish morphology over a half century.

Juola (2003) investigated samples from the magazine National Geographic over the period of 1939 to 2000. The material was divided into periods and for each sample in each period, linguistic distance to all other samples of the same period was calculated to assess the overall distance between two language samples. The measure has reportedly been applied successfully to the (synchronic) tasks of language identification, authorship analysis, language family identification and others (Juola, 2003: 81-83). This measure of distance was then correlated with the amount of time separating the samples. Subsequently, the rates of change over the different periods were compared to see if they were similar. This turned out not to be case: '1940s had less change (significantly so), than the 1970s, the 1970s had significantly less change than the 1950s and 1960s' (Juola, 2003:90), showing that rates of change can be uneven. Again, from a linguistic point of view, however, a number of difficulties are apparent in this study. Few and small samples from a single magazine can mean that synchronic variation (or breaks in editorial policy) might be responsible for at least some of the effects (cf. Millar, 2009; Leech, 2011). Most importantly, it is not clear what sorts of linguistic change were picked up – very obviously a critical point for linguistic analysis. While results appear plausible, the measurement of linguistic difference is essentially a black box, its inner workings, although mathematically clear, are not transparent in their impact on linguistic material. Some types of change might have influenced the measure more heavily than others and created a bias. As Juola himself points out, questions remain: 'Is [...] change primarily lexical [...]? Is the rate of lexical innovation different from the rate of syntactic innovation? Does this represent merely a pragmatic difference in what people choose to write/talk about, or is there a fundamental difference in the representation of language [...]?' (Juola, 2003:94). It is difficult to see how the procedure could help to answer these key questions.

In summary, previous research has supplied only impressionistic pointers to the likely speed of change in FL, but has demonstrated that there is very notable synchronic variation across text types. Research in the area of quantifying linguistic change generally has demonstrated that, for all the remaining difficulties, quantitative measurements of change are possible and, if derived in a reliable and transparent manner, offer insights quite unobtainable through other means, both in terms of the type of statements that can be made as well as generality and therefore validity which contrasts very sharply with extrapolations made on the basis of hand-picked example cases or lists.

The present study develops a reliable and transparent method to assess the speed of change in FL (and other constructs), relating it to baselines of synchronic variation and lexical change. The procedures, applied to a reference corpus of 20th century written German, supplies results on the specific question after the speed of FL change over the period, and also ascertains whether FL-density (i.e. the proportion of text that consists of FL) is subject to change over the period of observation, an interesting further aspect. As is demonstrated below, results address current shortcomings in research on FL and language change more generally, and in so doing substantially advance the understanding of both these areas.

3 Measuring Change in FL

3.1 Data

The data for the present study are taken from the Swiss Text Corpus (Bickel et al. 2009). This 20-million word corpus is a diachronic reference corpus of 20th century written German as used in Switzerland – one of only a few diachronic reference corpora suitable size for an investigation of this nature.[1] The corpus is topic balanced as well as balanced across four broad text types. The temporal structure is such that the corpus can be divided into five balanced sub-corpora covering consecutive 20-year periods. A summary is shown in table 1.

Table 1: Number of words in the Swiss Text Corpus (STC). *Note.* Number of words given in million, FI=fiction, JOU=journalistic texts, SUB=subject texts, FU=functional texts

	Period 1 (1900-1919)	Period 2 (1920-1939)	Period3 (1940-1959)	Period 4 (1960-1979)	Period 5 (1980-2000)	Total (1900-2000)
FI	0.7	1.0	0.9	0.9	0.7	4.2
JOU	0.4	0.3	0.7	0.7	1.0	3.2
SUB	1.2	1.2	1.4	1.5	1.5	6.7
FU	1.2	0.8	1.0	1.0	0.8	4.5
Total	3.3	3.3	4.0	4.0	4.0	18.6

[1] A similarly structured diachronic reference corpus is the 400-million-word Corpus of Historical American English (Davies, 2012), though this is only available in full-text with certain words removed (cf. https://www.corpusdata.org/limitations.asp); Google Books corpora in various languages are available only as n-gram lists and other diachonic corpora of sufficient size, such as the DWDS corpus (Geyken, 2007), tend to be publically available only via web-based search interfaces, making them difficult to work with.

For purposes of identification in corpus data, and in line with a characterisation of FL as phrases that are conventional pairings of a form and a unit of meaning in a speech community, FL was operationalised as in (1).

(1) n-grams of length 2 to 7 words that represent semantic units. Optionally, sequences can contain an internal slot extending over one word, so long as the sequence is attested in continuous form as well.

Semantic units were defined as word sequences possessing the sort of semantic unity typically found in words and structurally complete phrases. Semantic unity was also attributed to sequences that, while lacking this unity, can acquire it through the addition of a single, semantically or formally restricted slot (such as when *in search of* does not form a full semantic unit unless a slot on its right edge is added, i.e. *in search of X* where *X* is restricted semantically to something prized that is being pursued).

Items of FL were extracted automatically from corpus data using the extraction procedure developed in Buerki (2012). Briefly, this procedure consists of three main steps: in the first, n-grams of between 2 and 7 words in length are extracted using the N-Gram Processor (Buerki, 2014). In the second step, four main types of filter are applied: an additive stoplist eliminating sequences entirely composed of word forms from the top 200 most frequent words of German (as per the Leipzig Corpus Portal; anon, 2001), a frequency filter (eliminating n-grams below a frequency of four per million words), a document filter eliminating sequences that occur in fewer than three corpus documents, and a lexico-structural filter designed to remove sequences unlikely to be semantic units (e.g. sequences featuring the conjunction *dass* 'that' as the initial word of the sequence). In the final step of the procedure, the frequencies of n-grams of various lengths are consolidated such that shorter sequences that are included in longer sequences are not counted multiple times, and sequences that only occur as part of longer sequences are eliminated (cf. Buerki, 2017).

To assess the precision of this procedure, a random sample of 200 sequences automatically extracted from the data was assessed by two independent raters for compliance with the operationalisation. Their assessments agreed in 87% of cases (Cohen's kappa = 0.686) and shows that the procedure operates at a precision such that 71% of types and 77% of FL-tokens (average across raters) are operationalisation-compliant items of FL. The recall (i.e. completeness) of an extraction is more difficult to assess as the true number of operationalisation-compliant items of FL in the corpus is unknown. Notably, however, a minimum frequency of 4/M is far lower than that used in typical studies employing a frequency-based n-gram approach to FL (cf. Biber et al., 1999; Cortes, 2002) and a total of close to half a million FL-tokens were extracted per four million word sub-corpus. Extracted FL included functional formulae, collocations, multi-word units and other usual sequences. Table 2 shows examples of extracted sequences.

Table 2: Examples of extracted sequences.

Type	Example	Gloss
Formulae	*darüber hinaus*	what is more
	meines Erachtens	in my opinion
	zum Beispiel	for example

Collocations	*im Handel*	commercially available
	gesetzliche Grundlage	legal basis
	sehr empfohlen	highly recommended
Multi-word units	*immer noch*	Still
	ab sofort	starting immediately
	Klein- und Mittelbetriebe	small and medium-sized businesses
Other usual sequences	*stellt sich die Frage ob X*	begs the question whether X
	von Fall zu Fall	on a case by case basis
	weisst du noch	do you remember [when]

Notably, although the precision of the automatic extraction was manually assessed on a sample, no manual filtering of automatically extracted sequences took place and all extracted sequences were used in subsequent analyses. It is here argued that sequences that are not operationalisation-compliant are more than balanced out numerically by items of FL that could not be extracted and therefore the resulting number of extracted sequences remains on the conservative side of what is the likely number of FL-items contained in the data. Additionally, it was assumed that automatically extracted sequences on the whole do not behave in a manner that is so drastically different from the behaviour of the set of operationalisation-compliant FL that results are significantly affected. This assumption is warranted given the precision and good recall of the extraction procedure and the finding that length statistics for automatically extracted sequences and operationalisation-compliant items of FL in the data were closely similar. For these reasons, and for convenience, the term *FL* will subsequently be applied to the full set of automatically extracted sequences.

3.2 Method

To determine a possible change in FL-density across time, the number of word tokens that form part of FL and the number of word tokens *not* part of FL were compared across each of the five temporal sub-corpora of the STC using a chi-square test. Effect size was assessed using Cramer's V. FL-density for each period was also calculated using the formula used is shown in (2) to provide an accessible density metric.

$$\text{(2) FL-density} = \frac{\sum(\text{length of FL-item} \times \text{frequency})^{\text{FL-ITEM}_1\ \text{FL-ITEM}_2\ \ldots\ \text{FL-ITEM}_n}}{\text{number of word tokens in sub-corpus}}$$

This translates to multiplying the length in words of each extracted item of FL (FL-item 1 to n) by its frequency, then summing the resulting numbers. This is divided by the word count of each sub-corpus.

To measure extent of change across time, items of FL were extracted from each of the five temporal sub-corpora of the STC. In addition to lists of FL-items, word lists were also created for each sub-corpus used. In a first step toward establishing the extent of diachronic change, the similarity across sub-corpora representing different time periods was established (cf. figure 1). The degree of similarity was derived using a simple matching coefficient (cf. Oakes, 1998: 112-3), measuring the proportion of shared FL (and words) across two sub-corpora. The simple matching coefficient was calculated using (3) following Oakes (1998:112).

(3) $S_{SM} = \frac{m}{m+u} = \frac{m}{n}$

S_{SM} is similarity simple match, m is the number of matching items of FL (or words), u is the number of unmatched ones (i.e., those unique to each sub-corpus) and n is $m+u$, that is, the total number of items of FL (or words) in both texts. Since the sub-corpora compared were matched for size, simple similarity here provides a robust measure.

<insert figure 1 here>
Figure 1: Measuring change by establishing the shared proportion of FL-items across sub-corpora representing different time periods.

The measure was calculated for both type and token counts; for this purpose, the original measure, which is applied such that n represents the total number of types across both corpora compared and m represents the number of shared types, was modified as follows: to calculate similarity in types, n was taken as the mean number of types across the two sub-corpora compared (so that the result was the proportion of shared types in *one* sub-corpus which yields a more transparent metric for the purposes of this study); for tokens, m was taken as all instances of shared types (i.e. the total frequency of all shared types, regardless of whether their frequencies matched across corpora), and n represented the sum of tokens of both sub-corpora.

Using the proportion of shared FL and words across temporal sub-corpora in this manner is an elementary measure of similarity. More sophisticated similarity measures were considered, such as cosine similarity (Salton and McGill, 1983), but the simple matching coefficient is here preferred since its interpretation is straightforward, leading to maximally transparent and easily interpretable results. This is pivotal to avoiding the black box phenomenon discussed above and corresponds to a clear and plausible conceptualisation of similarity for the purposes of this application. As is demonstrated below, the simple matching coefficient performed well in detecting the type of change investigated, rendering more complex measures redundant.

<insert figure 2 here>
Figure 2: Pairwise measurements of similarity. *Note.* The flexibility of pairwise assessment of similarity enables measurements across adjacent time periods as well as further apart periods.

Pairwise comparison, in conjunction with the five time periods of the data, allowed for a flexible measuring of similarity across different periods as shown in figure 2. For example, similarity of FL across the century could be measured by comparing the list of FL-items of the first and the last period (shown as ② in figure 2), similarity across adjacent periods (or across any other span), by comparing their respective lists of FL-items (for example as shown in ①).

To establish the degree to which changes in similarity across temporal sub-corpora reflected diachronic change, two points of reference were established. The first was the proportion of shared FL or words (i.e. similarity) across contemporary texts. For this purpose, random sub-corpora were created by assigning contemporary corpus documents randomly to one of two sub-corpora of equal size and measuring similarity between those random sub-corpora. This provided a baseline for the

interpretation of similarity across temporal sub-corpora. A second point of reference was established by measuring similarity across the four different genres represented in the data. For this purpose, genre-specific sub-corpora were created and compared.

<insert figure 3 here>
Figure 3: Sub-corpora

An overview of comparisons made and sub-corpora used is given in figure 3. In total, the proportion of shared FL as well as of shared individual words was calculated across five time periods, four genres and two random halves. The same measure of similarity as per (3) was used for all comparisons. Various sizes of the sub-corpora were used, as will become apparent in the results section, below. The sub-corpus pairs being compared needed to be of equal size since similarity increases with corpus size (i.e., if sub-corpora of smaller size are compared, the proportion of shared words and FL will always be lower than if two sub-corpora of a larger size are compared). For the measure of similarity derived from one pair of sub-corpora to be comparable to other pairs, equal sub-corpus sizes across pairs compared were also required. Sub-corpora of appropriate size were created by assembling corpus documents with the required features into sub-corpora. If the resulting size was too large, a random subset of documents with the required features was used.

An overall measure of diachronic change was additionally devised based on these measures of similarity across sub-corpora. It expresses diachronic change in terms of the proportional reduction of similarity caused by diachronic change and is calculated using the formula in (4): Maximum similarity is taken as similarity across contemporary texts. From this figure of maximum similarity (S_{SM}synchronic), the similarity across temporally distant texts (i.e., S_{SM}diachronic; either across consecutive time periods or time periods with a time gap between them, cf. figure 2) is deducted, resulting in the reduction in similarity due to change. This resulting figure is then expressed as a proportion of the maximum similarity (i.e., it is divided by S_{SM}synchronic, the similarity across contemporary texts), yielding the proportional reduction in similarity due to diachronic change (DC). Since this last step, the relativisation on S_{SM}synchronic, yields a normalised measure, it should no longer depend on underlying sub-corpus size, but represents a measure comparable over different sub-corpus sizes despite the observation of dependency of similarity on corpus size.

$$(4)\quad DC = \frac{S_{SM}\text{synchronic} - S_{SM}\text{diachronic}}{S_{SM}\text{synchronic}}$$

For example, if the similarity measure for FL-tokens across two random sub-corpora came to a value of 0.7 (i.e., 70% of FL-tokens were shared), and the similarity measure for FL-tokens across the first and the last time period in the data (i.e., case ② in figure 2) came to 0.5 (i.e., a proportion of shared FL-tokens of 50%), the extent of diachronic change is calculated as shown in (5). The result (28.6%) represents the reduction in shared FL due to diachronic change.

$$(5)\quad DC = \frac{0.7 - 0.5}{0.7} = \frac{0.2}{0.7} = 0.286$$

Another way of thinking about this measure of change is of course in terms of the converse of similarity, that is, in terms of difference across sub-corpora. In this way, the difference across random sub-corpora represents synchronic variation (i.e., the

proportion of *non*-shared FL or words) and the difference across temporal sub-corpora represents both diachronic change *and* synchronic variation (since some of the difference across temporal sub-corpora is due to synchronic variation and some of it due to diachronic change). Deducting the difference across random sub-corpora from the difference across temporal sub-corpora thus yields the difference due to diachronic change. This is then relativised on the maximum possible similarity (i.e., the similarity across random sub-corpora, as before). This way of thinking can be expressed as DC' in (6), where 1-S_{SM} is the converse of similarity (i.e., the difference) across sub-corpora. DC' yields the same result as DC as demonstrated by (7), using the example figures from (5):

$$(6)\quad DC' = \frac{(1 - S_{SM}\text{diachronic}) - (1 - S_{SM}\text{synchronic})}{S_{SM}\text{synchronic}}$$

$$(7)\quad DC' = \frac{(1 - 0.5) - (1 - 0.7)}{0.7} = \frac{0.5 - 0.3}{0.7} = \frac{0.2}{0.7} = 0.286$$

The question after the speed of change was assessed by comparing the extent of change in FL across the five time periods with the extent of change in individual words across the same period of time, taking synchronic variation into consideration. In this way, both extent and speed of change could be quantified in a meaningful manner: types of change were clearly defined (change relating to words and FL) and the measure used was transparent and straightforward.

4 Results

Table 3: FL-density over time. *Note.* Corpus size normalised at 4 million words

Time period	1900-1919	1920-1939	1940-1959	1960-1979	1980-2000
FL-density	0.362	0.356	0.349	0.345	0.342

4.1 FL-density

Starting out with a consideration of FL-density across time, the numbers presented in table 3 show no dramatic change in densities. The chi-square test reveals, however, that time period and density are not independent of each other: χ^2 (4, N = 20,021,130) = 5220.999, $p < 0.0001$. Though statistically significant, the effect is minute (Cramer's V = 0.016). It is interesting to note that densities are perfectly negatively rank-ordered – they decrease steadily as time progresses.

Theoretical considerations led to an expectation that FL-density would not be subject to notable change. Counter to this, our data showed (for the first time as far as I am aware) that FL-densities are indeed subject to slight, but in this case consistent, changes across time. It will be interesting to see if the trend to slightly less formulaicity is robust in other data, or whether this is a feature of these particular data. Given the small effect size, however, it is not in doubt that FL plays a highly significant and constant role in language: the proportion of FL in language is not subject to large diachronic fluctuation, at least in the relatively short span of a century.

4.2 Extent and speed of change

Moving to results regarding extent and speed of change in FL, we consider the outcome of the various measures of diachronic change and synchronic variation. The large number of specific measurements taken (i.e., comparisons across various sub-corpora) means that not all of them can be narrated here. The full list of comparisons and their results are listed in the appendix where exact figures are given. In the following, selected results are presented primarily as graphs.

<insert figure 4 here>
Figure 4: Percentages of shared FL across random sub-corpora of various sizes.
Note. Corpus sizes apply to each of the sub-corpora compared pairwise.

As a first point of reference, we consider synchronic variation across different contemporary texts which serves as a baseline for assessments of change. Figure 4 shows the proportion of shared FL across random sub-corpora of different sizes (pairs of corpora compared at each level were of identical size). Results show that the proportion of shared FL increases in line with sub-corpus size. It is therefore vital for comparisons to be based on sub-corpora of identical size. The increase in the proportion of shared FL flattens out considerably as the shared proportion approaches the 90% mark (tokens) at sub-corpus size 9.6 million words. Due to the temporal structure of our data, the largest useful sub-corpus size (apart from the random sub-corpora shown in figure 4) is limited to the 3.3 million words of the smallest temporal sub-corpus (i.e., the size of the 1900-1919 time period, cf. table 1). At this sub-corpus size, there is no flattening out yet, meaning that proportions remain relative to corpus size used. As demonstrated below, this does not preclude precise measurements of change (especially if our DC measure is used), but merely means that sub-corpus size is a factor that needs to be considered. Figure 4 also clearly illustrates the importance, at any sub-corpus size, of considering synchronic variation when assessing proportions of shared FL diachronically: no two groups of texts are ever likely to share all their FL (or words). A final observation relating to the figures shown as well as subsequent figures presented is that similarity measured in types is somewhat lower than similarity measured in tokens. This indicates that shared types are of higher frequency than non-shared types, which is entirely plausible. Since types and tokens move in parallel here, the question of which of the two provides a more sensible measure does not pose itself urgently. For now, it is suggested that from the point of view of language in use, it is the token level which is more relevant since tokens represent actual linguistic items in use. We shall return to this consideration below.

<insert figure 5 here>
Figure 5: Percentages of shared FL across random sub-corpora in genres. *Note.*
Corpus size 0.8 million; FI=fiction, JOU=journalistic prose, SUB=subject texts, FU=functional texts, M=mean across the 4 genres, MIXED=mixed genres

Looking at synchronic variation across different random sub-corpora of contemporary texts in a genre by genre fashion (figure 5), it is clear that there are genre-specific differences: journalistic prose, for example, appears to be a more homogeneous genre than subject texts from the point of view of FL, the latter scoring the lowest proportion of shared FL across contemporary texts, the former scoring the

highest. Genre differences are significant (χ^2 (3, N = 1,002,255) = 445.29, p < 0.0001, figures for tokens), but again only a minute effect size is achieved (Cramer's V = 0.067). Results also show that synchronic variation does not differ greatly depending on whether shared FL are measured as mean across random sub-corpora of the same genre (labelled 'm') or across random sub-corpora of mixed genre (labelled 'mixed'). Importantly, this makes it unproblematic to use mixed-genre sub-corpora to assess synchronic variation.

Table 4: Percentages of shared FL across genres (S_{SM} multiplied by 100). *Note.* Figures based on sub-corpus sizes of 3.3 million words

		FI	FU	SUB	JOU	mean
FI	types	-	30.2	34.0	35.6	33.3
	tokens	-	46.8	51.4	53.3	50.5
FU	types	30.2	-	60.2	59.2	49.9
	tokens	46.8	-	75.5	74.4	65.7
SUB	types	34.0	60.2	-	63.9	52.7
	tokens	51.4	75.5	-	78.9	68.6
JOU	types	35.6	59.2	63.9	-	52.9
	tokens	53.3	74.7	78.9	-	69.0
mean	types	33.3	49.9	52.7	52.9	47.2
	tokens	50.5	65.7	68.6	69.0	63.4

Next, we consider synchronic variation of FL across genres rather than across random sub-corpora. Table 4 presents the results of pairwise comparisons across the four genre sub-corpora at sub-corpus size 3.3 million words. It is important again to remember that the exact proportions are dependent on sub-corpus size and are therefore specific to the size used. The general observations on relations among the proportions as drawn out immediately below, however, are not dependent on sub-corpus size and also held at sub-corpus size 0.8 million words (see appendix, entries 19 to 24, for the exact figures of those comparisons). The first general observation is that fiction is the most distinct genre in terms of FL, on average sharing only 33.3% of types and 50.5% of tokens with the other genres, whereas the other genres share around 50% of FL-types and between 65.7% and 69% of tokens with all the respective other genres on average at sub-corpus size 3.3 million. This may be due, in part, to the inclusion of dialogue in some works of fiction. As mentioned above, previous research indicates that FL are particularly sensitive to the spoken/written division. The most closely similar genres were subject texts and journalistic prose. Again, this is plausible – the corpus compilers themselves independently remarked on the closeness of these two genres (cf. Bickel et al., 2009:13). As the point of reference for comparison with diachronic FL-change, the overall mean of the shared proportion of FL across genres (the figures in the lower right corner of table 4) will be used. It should be kept in mind that this figure masks the patterning of three relatively similar genres against one rather different genre (fiction). If only the three genres of functional texts (FU), subject texts (SUB) and journalistic prose (JOU) were considered, the mean shared proportion of FL at sub-corpus size 3.3 million would come to 61.1% for types and 76.4% for tokens – markedly higher than if fiction is included, but still below the figures for similarity across random sub-corpora, thus replicating earlier findings regarding the distinctiveness of genres in FL-terms.

4.2.1 Diachronic change in words and FL

Moving on to diachronic comparisons, figure 6 shows change as measured by the proportion of shared FL across various temporally ordered sub-corpora. Moving from left to right, this is the proportion of shared FL across random sub-corpora (for types and tokens respectively), followed by shared FL across time period 5 (1980-2000) and period 4 (1960-1979). The third pair of bars is across period 5 (1980-2000) and period 3 (1940-1959), the fourth across period 5 (1980-2000) and period 2 (1920-1939), the fifth across period 5 (1980-2000) and period 1 (1900-1919). Finally, the mean proportion of shared FL across genres (as per the lower right corner of table 4) is added for comparison.

<insert figure 6 here>
Figure 6: Percentages of shared FL across time at 3.3 million. *Note.* Random = comparison across random sub-corpora; p5 = 1980-2000, p4 = 1960-1979, p3 = 1940-1959, p2 = 1920-1939, p1 = 1900-1919; m genres = mean shared FL across genres.

Looking at figure 6, we can make a number of key observations. First, as would be expected if the method takes sensible measurements, the proportion of shared FL across random sub-corpora is the highest. Strikingly, however, the other contemporary comparison, that across genres, shows the lowest proportion of shared FL, lower even than that across the first and the last period of the twentieth century (p5:1). Thus the mean variation across genres is greater than the extent of change across the century. The other key observation is that, again as one would expect if the measure behaves plausibly, the proportion of shared FL drops as time periods at greater distance from each other are compared: adjacent time periods (p5:4) show the highest proportion, those with an intervening 20 years (p5:3) a lower proportion, those with an intervening 40 years (p5:2) lower again and those with an intervening 60 years (p5:1) the lowest proportion of shared FL (i.e., a perfect rank-correlation). This indicates that 1) change in FL over a century (and indeed over adjacent periods covering 40 years back to back) can be detected, and 2) that the measure employed is well suited for detecting change over time. Proportions for tokens are higher than those for types throughout, simply indicating that more frequent types are more likely to be shared ones, again as might be expected. Looking at the proportions labelled p5:4 to p5:1, the impression of a slowing pace of change is gained. This is a product, however, of the manner of comparison (period 5 being compared with each of the remaining periods) which means that unequal stretches of time are being compared. Since change is not necessarily linear, an assessment of the pace of change during various sub-periods of time needs to be based on comparisons of change across equal lengths of time. The results of such a comparison will be presented below. Before that, however, we verify the basic robustness of the patterning in figure 6.

<include figure 7 here>
Figure 7: Percentages of shared FL across time (0.8 M). *Note.* Random = comparison across random sub-corpora; p5 = 1980-2000, p4 = 1960-1979, p3 = 1940-1959, p2 = 1920-1939, p1 = 1900-1919; m genres = mean shared FL across genres

For this purpose, the same comparisons as in figure 6 were conducted on the basis of a sub-corpus size of 0.8 million words (figure 7), and also using sub-corpora of only a

single genre (subject texts). The figures for subject texts were virtually identical to those of figure 7 and showed that mixed-genre texts behave in closely similar ways to single-genre texts (cf. appendix, entries 37-40 and 41-44, for details). While, for the reasons indicated above, the percentages of shared FL are not comparable across different corpus sizes (easily recognised by the lower actual percentages), figure 7 demonstrates that the pattern of relations is entirely robust: again, the proportion of shared FL across random sub-corpora is the highest and that across genres the lowest. Again also, proportions of shared FL drop as time periods at greater distance from each other are compared.

Table 5: Diachronic change in FL across adjacent time periods (3.3 M). *Note.* Figures calculated using formula in (4); p1 = 1900-1919, p2 = 1920-1939, p3 = 1940-1959, p4 = 1960-1979, p5 = 1980-2000

	p1:2	p2:3	p3:4	p4:5
types	0.036	0.073	0.071	0.086
tokens	0.018	0.040	0.038	0.053

Looking at whether FL changed at a constant pace or showed periods of slower and faster change, table 5 presents diachronic change (DC) across adjacent time periods (p1:2, p2:3, p3:4, p4:5), calculated using the formula in (4) above, based on data derived from corpus size 3.3 million. Figures vary between 3.6% and 8.6% for types and 1.8% and 5.3% for tokens. The smallest amount of change is measured across periods 1 and 2 (1900-1919 and 1920-1939) and the greatest change across periods 4 and 5 (1960-1979 and 1980-2000), with the remaining time periods occupying the middle ground. This shows a trend toward an accelerating pace of change, though not entirely consistently so (across periods 2 and 3, there is more change than across periods 3 and 4).

<insert figure 8 here>
Figure 8: Percentages of shared individual words across time at 3.3 million. *Note.* Random = comparison across random sub-corpora; p5 = 1980-2000, p4 = 1960-1979, p3 = 1940-1959, p2 = 1920-1939, p1 = 1900-1919; m genres = mean shared words across genres

<insert figure 9 here>
Figure 9: Change in FL and words at sub-corpus size 3.3 million. *Note.* Random = comparison across random sub-corpora; p5 = 1980-2000, p4 = 1960-1979, p3 = 1940-1959, p2 = 1920-1939, p1 = 1900-1919; m genres = mean shared FL/words across genres

Moving on to change in individual words, figure 8 shows the same comparisons as figure 6 but this time for individual words. The difference between the values for types and tokens is far more pronounced than in FL, showing that fewer types are shared, but their frequency is higher. Crucially, however, indications of lexical change, in contrast to FL-change, are not as easily detectable over the period covered by the data: word tokens across random sub-corpora, the various temporal sub-corpora and genre sub-corpora are almost identical. They do show minute differences consistent with the pattern found in FL-change in that the figure for *random* is

slightly higher than that for the temporal sub-corpora and the similarity of temporal sub-corpora reduces ever so slightly in line with the increasing time gaps of the comparisons (exact figures listed in the appendix). Word types show a more readily recognisable pattern, again similar to that found in FL but the differences are much slighter. For both types and tokens, the mean proportion of shared words across genres is higher than that across p1 (1900-1919) and p5 (1980-2000), indicating that the extent of change in words over the century, albeit barely detectable, is greater than the variation across genres which was not the case for FL. It was argued above that tokens are more relevant than types. This can be illustrated using the results in figure 8: although only 44% of word types are shared across random sub-corpora, these are so highly frequent that they make up 94% of word forms in use. Conversely, the 56% of word types that are not shared account for only 6% of word forms in actual use. It is therefore argued that, particularly in cases where type and token figures diverge notably, it is the token figures that supply the most accurate information on language use.

If the values for FL-tokens from figure 6 are compared to word-tokens from figure 8, the differences between lexical change and FL-change become starkly apparent. Figure 9 illustrates the comparison: shared word tokens reduce only by 2 percentage points between random sub-corpora and p5:1 (the full extent of temporal difference), whereas shared FL reduce by 14 percentage points.

To test if these differences between words and FL might be due to the minimum frequency threshold for FL (four occurrences per million words) employed during extraction, frequent words (i.e., words appearing with minimum frequency of 4/M) were also tested. Results show that the enforcement of a minimum word frequency of 4 occurrences per million words (4/M) does not change the picture decisively (cf. figure 10). Change in frequent words was more clearly detectable than change in all word types, but crucially, change was still of notably smaller extent than change in FL. Consequently, the differences between words and FL in terms of diachonic change are not caused by the minimum frequency requirement of FL, although frequent words do behave slightly more similarly to FL than do all words, particularly when looking at word types.[2]

<insert figure 10 here>
Figure 10: Change in FL and frequent words at sub-corpus size 3.3 million. *Note.* Frequent words are words occurring at least four times per a million words of running text; random = comparison across random sub-corpora; p5 = 1980-2000, p4 = 1960-1979, p3 = 1940-1959, p2 = 1920-1939, p1 = 1900-1919; m genres = mean shared FL/words across genres.

4.2.2 Extent and speed of change

Having reviewed proportions across random sub-corpora, various temporal sub-corpora and across genres, we are now in a position to assess and compare the extent of change in FL and words over the 20th century. The extent of change was calculated using the measure of diachronic change (DC) introduced in (4) above, that is, it was

[2] An interesting facet of this closer similarity between frequent words and FL is also that mean variation across genres, which is smaller than the full extent of change over the century in all words, shifts to become greater than the full extent of change over the century when looking at frequent words (cf. entries 25-36 of the appendix).

measured in terms of the reduction in shared FL-items due to change. Frequent words were included in the comparison to demonstrate that they pattern very similarly to words in general.

Table 6: Extent of change in FL, words and frequent words across the century based on sub-corpus size 3.3 million. *Note.* Numbers represent the proportional reduction in similarity across contemporary texts due to diachronic change. Calculated using the formula in (4)

	Types	Tokens
Formulaic language	0.250	0.167
Words	0.193	0.021
Frequent words	0.174	0.028

Results are shown in table 6. For both types and tokens, FL-change was greatest in extent. For types, temporal distance across the whole period of investigation (p5:1) resulted in a 25% lower proportion of shared FL-types, compared to contemporary texts. For tokens, the figure was 16.7%. For words as well as frequent words, these figures were much lower: for types, temporal distance reduced the proportion of shared words by less than 20%, for tokens, it was less than 3%. The contrast between FL and words varies notably between types and tokens, although the general picture that FL-change is much faster holds from both points of view. It was suggested above that from the point of view of language in use, the token level is far more relevant. This is again demonstrated by the figures in table 6: the 19.3% of change in word-types represent only 2.1% of change in word tokens as they appear in actual language use, meaning that most of the word tokens encountered in texts remain unchanged. While reliably detectable in the data, this is fairly minor. Though the extent of change in FL-types is also greater than change in terms of tokens, the 25% change in FL-types still represents a change of nearly 17% in terms of actual FL-tokens in texts. This is more than seven times the change in word tokens (2.1%). Plainly, FL changed to a far greater extent over the period of observation than words (whether all words or only frequent words are considered).

It is now possible to summarise findings on the extent of change in FL: this change was notably larger than lexical change over the same time period. This translates to a faster speed of change for FL. In case of tokens, FL-change was multiple times as fast as lexical change. For types, argued to be a less relevant measure, the speed was almost a third faster.[3]

5 Discussion

This paper investigated diachronic change in FL from a bird's eye view. It focussed on a broad quantification of change, as comprehensive as possible with respect to the source data. This avoids the impressionistic nature of generalisations made on the basis of very few selected examples. Instead, a reliable quantification of change over large amounts of data – in our case over the entire data contained in the STC – is achieved. Downsides of a comprehensive quantification of change, as pointed out above, can arise when procedures are employed that are of a complexity that unduly

[3] Taking 19.3% as a base, 25% is 29.5% faster.

complicates the interpretation of results, or even renders a precise interpretation impossible. Problems also arise when it remains unclear what sort of change is being measured, such as when overall measures of difference are applied. Both of these potential downsides are avoided in the analyses shown: the measures employed (similarity across sub-corpora in terms of the proportion of shared types and tokens and the newly developed measure of DC as the proportional reduction in synchronic similarity due to diachronic change) are maximally transparent and therefore easily interpretable while retaining their effectiveness and robustness as measures of change, in the case of DC even across varying corpus sizes. The type of linguistic change measured is equally clearly delineated as pertaining to FL on the one hand, and individual words (as well as frequent words) on the other.

Regarding the first question pursued, it was expected that FL-density would remain stable across the period of investigation. Results showed that FL-density indeed remained near a third of running words being part of FL, though very small and consistent differences across time are detectable, indicating that the degree of formulaicity can vary somewhat with time – an intriguing finding that will benefit from further research.

Regarding the question after the speed of change in FL, analyses revealed that FL-change proceeds very much faster than lexical change over the period. If measured using what was argued to be the most relevant unit of tokens, the speed of FL-change is a multiple of the speed of lexical change.[4] This is a remarkable finding for the field of historical linguistics: hitherto the fastest type of change has generally been held to be lexical change (cf. Algeo 1980:264; Trask and Millar 2010:7). The results presented show this to be inaccurate. It was shown not only that FL-change is much faster than lexical change, but also that the two show such distinctly different patterns of change that a distinction between the two must be made: FL-change appears not to be a (special) case of lexical change, but a different type of change altogether.

The rate of FL-change was further found to vary over the period of investigation. This is consistent with findings by Juola (2003) and indeed more widely (cf. Bauer 1994), although our findings cannot confirm that population size is a factor in the speed of change (Nettle 1999) as both the population as well as the rate of change increased over the period of investigation. Again, this opens interesting avenues for further research, for example on possible causes of increases and decreases of rates of change in FL.

It is furthermore notable that the full extent of FL-change over the period of investigation remains slightly smaller than the average extent of synchronic variation across different genres, indicating a high sensitivity of the FL in our data to genre. This agrees with previous research that has robustly demonstrated FL to be highly sensitive to genre differences, and since genres are cultural constructs, to the cultural context (cf. e.g. Kuiper, 2009: 17).

It could be asked whether the rapidity of FL-change might be largely due to technical or mechanical rather than linguistic reasons: it may be hypothesized, for example, that the comparatively high speed of FL-change may be merely due to FL consisting of larger units of language than words which are therefore computationally bound to display less similarity across texts. However, the faster speed of FL-change is not due to a general tendency of words to be more similar across sub-corpora than

[4] In terms of types, FL-change is about a third faster than lexical change.

FL, since synchronic variation was controlled for (although words do indeed show a somewhat higher degree of similarity across random sub-corpora), cf. figure 9. Further, unit size itself cannot serve as an explanation for results shown: syntactic structures (similarly larger units of language than words) are thought to change much more slowly than words whereas some cases of sound change (notably involving very small units) appear to proceed fairly rapidly with differences readily detectable between generations of speakers (cf. Labov, 1972). [5] Altintas et al. (2007) also showed that extensive morphological change, again involving items smaller than words, can occur over brief periods. It seems clear, therefore, that the comparative rapidity is a feature of FL-change that warrants investigation and explanation from a linguistic point of view.

The findings of this study significantly advance the current state of research in two important fields. For the study of FL, they provide a solid assessment, based on a general quantification, of the speed of change in FL. It is confirmed that FL changes at a rapid pace – at a multiple of the rate of lexical change. This, along with the finding that FL-density remains generally fairly stable across time, even while the speed of FL-change appears variable, adds fascinating new aspects to the study of FL, an area increasingly recognised as central to the workings of language at large. One of the many possible implications of these findings to be explored is that the connection between FL and its cultural context is so close that a rapid rate of change is needed for FL to fulfil its functions: if the social and cultural changes of the twentieth century as reflected in the data used for this investigation had passed FL by without notable change, it would be difficult to argue for a close link between FL and a changing social and cultural context. Findings also confirm that FL is highly genre-sensitive, notably in contrast to words.

For the study of language change in general, this study shows that lexical change – hitherto thought to be the fastest type of linguistic change – is in fact notably slower than change in FL. Beside demonstrating a method to validly and robustly determine relative speed of change, the study has also introduced an easily interpretable new, independent measure of the speed of change, potentially applicable to any linguistic construct that can be counted in corpus data.

Funding

The work was supported in part by the Swiss National Science Foundation under Grant 118724.

References

Ädel, A. and B. Erman. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.

[5] It might be the (partial) abstraction away from lexically specific forms that allows certain constructions to remain far more constant across time than constructions at the lexically filled level (i.e. FL), but research clearly is insufficient to draw conclusions at this stage. More robust quantifications of syntactic change in future work might yet show the speed of syntactic change to have been assessed too conservatively in currently available research.

Algeo, J. (1980). Where do all the new words come from? *American Speech*, 55(4), 264-277.
Allerton, D.J. (1984). Three (or four) levels of word cooccurence restriction. *Lingua*, 63(1), 17-40.
Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. P. Cowie (ed.). *Phraseology: theory, analysis and applications.* Oxford: Oxford University Press. 101-22.
Altintas, K., F. Can and J.M. Patton. (2007). Language change quantification using time-separated parallel translations. *Literary and Linguistic Computing*, 22(4), 375-93.
Anon. (2001). *Wortlisten* [Data file]. http://wortschatz.uni-leipzig.de/html/wliste.html [accessed 13 July 2009].
Bally, C. (1909). *Traité de stylistique française, premier volume.* Paris: Librairie C. Klincksieck.
Bauer, L. (1994). *Watching English change: an introduction to the study of linguistic change in standard Englishes in the twentieth century.* London: Longman.
Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
Biber, D. (2006). *University language: a corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
Biber, D. and F. Barbieri. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-86.
Biber, D. (2009). A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275-311.
Biber, D., S. Conrad and V. Cortes. (2004). If you look at …: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
Biber, D., S. Conrad, and V. Cortes. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson and A. M. McEnery (eds). *Corpus Linguistics by the Lune: a festschrift for Geoffrey Leech*, 71-92.
Bickel, H., M. Gasser, A. Häcki Buhofer, L. Hofer and Ch. Schön. (2009). Schweizer Text Korpus - theoretische Grundlagen, Korpusdesign und Abfragemöglichkeiten. *Linguistik Online*, 39(3), 5-31.
Bischof, B.B. (2008). *Französische Kollokationen diachron: Eine korpus-basierte Analyse*. Stuttgart: Universität Stuttgart.
Boers, F., J. Eyckmans, J. Kappel, H. Stengers and M. Demecheleer. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245-61.
Bromham, L., X. Hua, T.G. Fitzpatrick and S.J. Greenhill. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences*, 112(7), 2097–2102.
Bubenhofer, N. (2009). *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin: De Gruyter.
Buerki, A. (2017). Frequency Consolidation among word n-grams: A practical Procedure. In R. Mitkov (ed.). *Computational and Corpus-Based Phraseology.* Berlin: Springer. 432-446.
Buerki, A. (2016). Formulaic sequences: a drop in the ocean of constructions or something more significant? *European Journal of English Studies*, 20(1), 15-34.
Buerki, A. (2014). *N-gram processor 0.4.* [software] http://buerki.github.io/ngramprocessor/ [accessed 13 November 2014]

Buerki, A. (2012). Korpusgeleitete extraktion von Mehrwortsequenzen aus (diachronen) Korpora. In N. Filatkina, A. Kleine-Engel, M. Dräger, und H. Burger (eds), *Aspekte der historischen Phraseologie und Phraseographie* (p. 263-92). Heidelberg: Universitätsverlag Winter.

Burger, H., D. Dobrovol'skij, P. Kühn and N.R. Norrick. (2007). Phraseology: subject area, terminology and research topics. In H. Burger, D. Dobrovol'skij, P. Kühn and N. R. Norrick (eds). *Phraseology: an international handbook of contemporary research.* Berlin: de Gruyter. 11-19.

Burger, H., A. Häcki Buhofer and A. Sialm. (1982). *Handbuch der Phraseologie.* Berlin: de Gruyter.

Burger, H. and A. Buhofer. (1981). Phraseologie als Indikator für Text- und Stiltypen. *Wirkendes Wort*, 6, 377-98.

Burger, H. and A. Linke. (1998). Historische Phraseologie. In Besch, W., A. Betten, O. Reichmann and S. Sonderegger (eds). *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Berlin: de Gruyter. 743-55.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge: Cambridge University Press.

Bynon, T. (1977). *Historical linguistics*. Cambridge: Cambridge University Press.

Chesley, P. and R.H. Baayen. (2010). Predicting new words from newer words: lexical borrowings in French. *Linguistics*, 48(6), 1343-74.

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice and D. Biber (eds). *Using Corpora to Explore Linguistic Variation.* Amsterdam: John Benjamins. 131-146.

Crowley, T. and C. Bowern. (2010). *An introduction to historical linguistics (4th ed.).* Oxford: Oxford University Press.

Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics*, 3(3-4), 239-66.

Dabrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, 25(4), 617-654.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word corpus of historical American English. *Corpora*, 7(2), 121-157.

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25-53.

Erman, B. and B. Warren. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.

Feilke, H. (1994). *Common sense-Kompetenz: Überlegungen zu einer Theorie des "sympathischen" und "natürlichen" Meinens und Verstehens.* Frankfurt am Main: Suhrkamp.

Feilke, H. (2003). Textroutine, Textsemantik und sprachliches Wissen. In A. Linke, H. Ortner and P. Portmann-Tselikas (eds). *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis.* Tübingen: Niemeyer. 209-30.

Fillmore, C.J., P. Kay and M.C. O'Connor. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language*, 64(3), 501-38.

Fodor, I. (1965). *The rate of linguistic change.* The Hague: Mouton.

Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (ed.). *Idioms and collocations: corpus-based linguistic and lexicographic studies.* London: Continuum. 23-40.

Greenhill, S. J., X. Hua, C.F. Welsh, H. Schneemann, and L. Bromham (2018). Population size and the rate of language evolution: A test across Indo-European, Austronesian, and Bantu languages. *Frontiers in Psychology*, 9, 576.
Gries, S. (2010). Bigrams in registers, domains and varieties: a bigram gravity approach to the homogeneity of corpora. In *Proceedings of Corpus Linguistics 2009*. http://ucrel.lancs.ac.uk/publications/ cl2009/404 FullPaper.doc [accessed 13 October 2012].
Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
Hyland, K., and Jiang, F. K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383–407.
Jespersen, J.O.H. (1904). *How to teach a foreign language*. London: Allen & Unwin.
Johnson, L. (1976). A Rate of change index for language. *Language in Society*, 5(2), 165-72.
Juola, P. (2003). The time course of language change. *Computers and the Humanities*, 37(1), 77-96.
Juola, P. (2005). Language change and historical inquiry. In *Proceedings of the XVI international conference of the Association for History and Computing (AHC 2005).* Amsterdam: Royal Netherlands Academy of Arts and Sciences. 169-75.
Juola, P. (2007). Authorship attribution. *Foundations and Trends in Information Retrieval*, 1(3), 233-334.
Juola, P. and R.H. Baayen. (2005). A controlled-corpus experiment in authorship identification by cross-entropy. *Literary and Linguistic Computing*, 20, 59-67.
Kopaczyk, J. (2012). Applications of the lexical bundles method in historical corpus research. In P. Pezik (ed.). *Corpus data across languages and disciplines*. 83-95. Berlin: Peter Lang.
Kuiper, K. (2009). *Formulaic genres*. Houndmills: Palgrave Macmillan.
Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
Langacker, R. W. (2008). *Cognitive grammar: a basic introduction*. Oxford: Oxford University Press.
Leech, G. N. (2011). The modals ARE declining: Reply to Neil Millar's "modal verbs in TIME: Frequency changes 1923–2006", International Journal of Corpus Linguistics 14:2 (2009), 191–220. *International Journal of Corpus Linguistics*, 16(4), 547-564.
Lieven, E. and S. Brandt. (2011). The constructivist approach. *Infancia y aprendizaje*, 34(3), 281-296.
Mair, C. (2006). *Twentieth-century English: history, variation and standardization.* Cambridge: Cambridge University Press.
Millar, N. (2009). Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics*, 14(2), 191-220.
Moran, S., D. McCloy and R. Wright. (2012). Revisiting population size vs. phoneme inventory size. *Language*, 88 (4), 877–893.
Myles, F. (2004, 8). From data to theory: the over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society*, 102(2), 139-168.
Nattinger, J.R. and J.S. DeCarrico. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.
Nettle, D. (1999). Is the rate of linguistic change constant? *Lingua*, 108(2-3), 119-136.

Oakes, M.P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
Pawley, A. and F. Syder. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (eds). *Language and communication.* Harlow: Longman. 191-226.
Pawley, A. (2001). Phraseology, linguistics and the dictionary. *International journal of lexicography*, 14(2), 122-34.
Pei, M. (1952). The story of English. Philadelphia: Lippincott.
Salton, G. and M.J. McGill. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
Sandøy, H. (2009). Quantifying linguistic changes experiments in Norwegian language history. In *Historical linguistics 2007: Selected papers from the 18th international conference on historical linguistics*, Montreal, 6-11 August 2007. Vol. 308:285.
Sankoff, D. (1970). On the rate of replacement of word-meaning relationships. *Language*, 46(3), 564.
Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.
Sorhus, H.B. (1977). To hear ourselves – implications for teaching English as a second language. *English Language Teaching Journal*, 31(3), 211-21.
Stubbs, M. (2002). *Words and phrases: corpus studies of lexical semantics.* Oxford: Blackwell.
Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21(2), 121-37.
Swadesh, M. (1959). Linguistics as an instrument of prehistory. *South-Western Journal of Anthropology*, 15(1), 20-35.
Trask, R.L. and R.M. Millar. (2010). *Why do languages change?* Cambridge: Cambridge University Press.
Van Lancker-Sidtis, D. and G. Rallon. (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language and Communication*, 24(3), 207-40.
Wichmann, S. and E.W. Holman. (2009). Population size and rates of language change. *Human Biology*, 81(3), 259-274.
Wichmann, S. (2008). The emerging field of language dynamics. *Language and Linguistics Compass*, 2 (3), 442-455.
Wray, A. and M.R. Perkins. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, 20(1), 1-28.
Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.
Wray, A. (2008). *Formulaic language: pushing the boundaries.* Oxford: Oxford University Press.

Appendix

Comparisons across random sub-corpora

	sub-corpus size	time period	genre	comparison type	shared FL-types / mean FL-types/S_{SM}	shared FL-tokens / sum of FL-tokens/S_{SM}
1	0.8 M	p4+5	FI	FL	7516 / 15022 / 0.500	201673 / 286532 / 0.704
2	0.8 M	p4+5	JOU	FL	7242 / 14172 / 0.511	179334 / 255272 /0.703
3	0.8 M	p4+5	SUB	FL	5217 /11415 / 0.457	143237 / 223207 / 0.642
4	0.8 M	p4+5	FU	FL	5768 / 12519 / 0.461	151720 / 237244 / 0.64

5	0.8 M	p4+5	mixed	FL	6068 / 12695 / 0.478	157856 / 238214 / 0.663
6	1.6 M	mixed	mixed	FL	8135 / 14406 / 0.565	368482 / 499319 / 0.738
7	3.3 M	p4+5	mixed	FL	7939 / 11438 / 0.694	748183 / 906656 / 0.825
8	6.4 M	mixed	mixed	FL	10999 / 14647 / 0.751	1709398 / 1971102 / 0.867
9	9.6 M	mixed	mixed	FL	11346 / 14503 / 0.782	2598222 / 2929893 / 0.887
10	3.3 M	p4+5	mixed	words	90034 / 205960 / 0.437	6177167 / 6546696 / 0.944
11	3.3 M	mixed	mixed	freq.words	12894 / 16909 / 0.763	5558403 / 5728150 / 0.97
12	0.8 M	p4+5	mixed	words	33184 / 82227 / 0.404	1454075 / 1610653 / 0.903

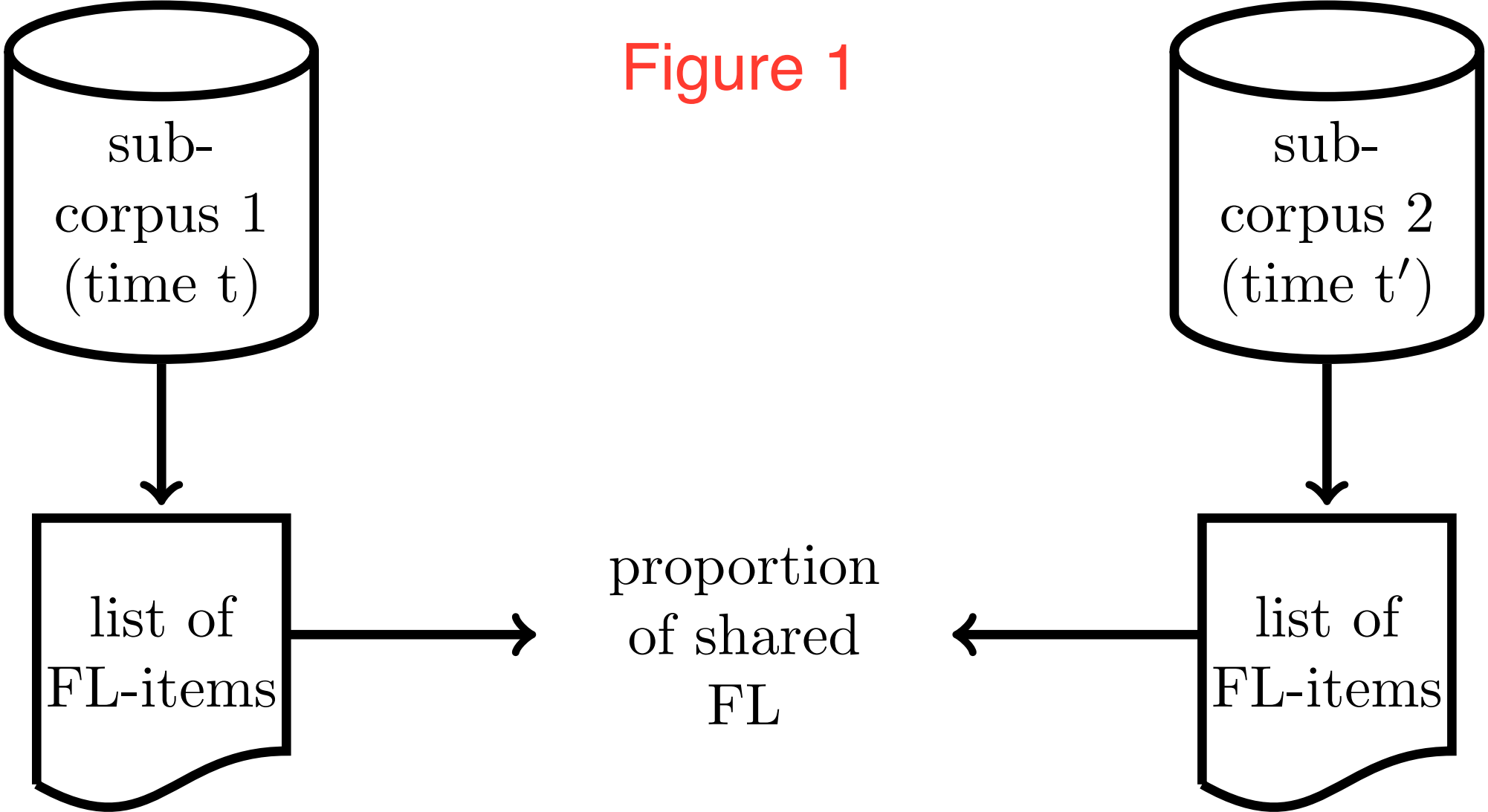
Comparisons across genre groups

	sub-corpus size	time period	comp.-type	comparison across	shared types / mean types / S_{SM}	shared tokens / sum of tokens / S_{SM}
13	3.3 M	mixed	FL	FI-FU	3833 / 12712 / 0.302	493170 / 1054044 / 0.468
14	3.3 M	mixed	FL	FI-SUB	4298 / 12641 / 0.340	537953 / 1046868 / 0.514
15	3.3 M	mixed	FL	FI-JOU	4643 / 13051 / 0.356	569320 / 1067598 / 0.533
16	3.3 M	mixed	FL	FU-SUB	6902 / 11456 / 0.602	707460 / 937508 / 0.755
17	3.3 M	mixed	FL	FU-JOU	7019 / 11866 / 0.592	715909 / 958238 / 0.747
18	3.3 M	mixed	FL	SUB-JOU	7533 / 11795 / 0.639	750810 / 951062 / 0.789
19	0.8 M	p4+5	FL	FI-FU	3256 / 13631 / 0.239	103299 / 260198 / 0.397
20	0.8 M	p4+5	FL	FI-SUB	3774 / 13327 / 0.283	114952 / 255272 / 0.450
21	0.8 M	p4+5	FL	FI-JOU	4060 / 14720 / 0.276	123620 / 272607 / 0.454
22	0.8 M	p4+5	FL	FU-SUB	4984 / 11880 / 0.420	136192 / 228880 / 0.595
23	0.8 M	p4+5	FL	FU-JOU	5763 / 13272 / 0.434	152056 / 246215 / 0.618
24	0.8 M	p4+5	FL	SUB-JOU	5669 / 12969 / 0.437	149681 / 241289 / 0.620
25	3.3 M	mixed	words	FI-FU	59113 / 175160 / 0.337	6027993 / 6545691 / 0.921
26	3.3 M	mixed	words	FI-SUB	61671 / 176753 / 0.349	6066969 / 6547175 / 0.927
27	3.3 M	mixed	words	FI-JOU	63456 / 175615 / 0.361	6112144 / 6552988 / 0.933
28	3.3 M	mixed	words	FU-SUB	82007 / 207582 / 0.395	6131425 / 6581998 / 0.932
29	3.3 M	mixed	words	FU-JOU	83840 / 206444 / 0.406	6166037 / 6587811 / 0.936
30	3.3 M	mixed	words	SUB-JOU	84159 / 208037 / 0.405	6173391 / 6589295 / 0.937
31	3.3 M	mixed	freq.words	FI-FU	7781 / 15697 / 0.496	5296790 / 5843519 / 0.906
32	3.3 M	mixed	freq.words	FI-SUB	7992 / 15468 / 0.517	5352192 / 5841011 / 0.916
33	3.3 M	mixed	freq.words	FI-JOU	8198 / 15425 / 0.531	5407843 / 5855252 / 0.924
34	3.3 M	mixed	freq.words	FU-SUB	11906 / 17195 / 0.692	5494957 / 5759092 / 0.954
35	3.3 M	mixed	freq.words	FU-JOU	12041 / 17152 / 0.702	5534580 / 5773333 / 0.959
36	3.3 M	mixed	freq.words	SUB-JOU	12360 / 16923 / 0.730	5565057 / 5770825 / 0.964

Comparisons across temporal sub-corpora

	sub-corpus size	genre	comp.-type	comparison across	shared types / mean types / S_{SM}	shared tokens / sum of tokens / S_{SM}
37	0.8 M	mixed	FL	p5:4	5587 / 12483 / 0.448	151479 / 237269 / 0.638
38	0.8 M	mixed	FL	p5:3	5234 / 12788 / 0.409	145655 / 243069 / 0.599
39	0.8 M	mixed	FL	p5:2	5093 / 12745 / 0.400	144753 / 247513 / 0.585
40	0.8 M	mixed	FL	p5:1	4855 / 12747 / 0.381	139385 / 246054 / 0.567
41	0.8 M	SUB	FL	p5:4	5298 / 12200 / 0.434	144017 / 233507 / 0.617
42	0.8 M	SUB	FL	p5:3	5418 / 13012 / 0.416	148756 / 246108 / 0.604
43	0.8 M	SUB	FL	p5:2	4993 / 12376 / 0.403	140712 / 238460 / 0.59
44	0.8 M	SUB	FL	p5:1	4889 / 12695 / 0.385	141965 / 245670 / 0.578
45	3.3 M	mixed	FL	p5:4	7343 / 11580 / 0.634	719150 / 920662 / 0.781
46	3.3 M	mixed	FL	p5:3	6602 / 11597 / 0.569	674212 / 920479 / 0.733
47	3.3 M	mixed	FL	p5:2	6253 / 11674 / 0.536	658231 / 936492 / 0.703
48	3.3 M	mixed	FL	p5:1	6187 / 11879 / 0.521	652162 / 948621 / 0.688
49	3.3 M	mixed	FL	p4:3	7472 / 11587 / 0.645	738229 / 930167 / 0.794
50	3.3 M	mixed	FL	p3:2	7513 / 11681 / 0.643	749739 / 945997 / 0.793
51	3.3 M	mixed	FL	p2:1	8001 / 11962 / 0.669	789773 / 974139 / 0.811
52	3.3 M	mixed	words	p5:4	83575 / 202692 / 0.412	6141893 / 6549222 / 0.938
53	3.3 M	mixed	words	p5:3	77095 / 200735 / 0.384	6078976 / 6531566 / 0.931
54	3.3 M	mixed	words	p5:2	71598 / 195947 / 0.365	6051468 / 6522543 / 0.928
55	3.3 M	mixed	words	p5:1	68788 / 194884 / 0.353	6026653 / 6522723 / 0.924
56	3.3 M	mixed	freq.words	p5:4	12178 / 16853 / 0.723	5523481 / 5739157 / 0.962
57	3.3 M	mixed	freq.words	p5:3	11294 / 16904 / 0.668	5459032 / 5729950 / 0.95.3
58	3.3 M	mixed	freq.words	p5:2	10684 / 16648 / 0.642	5435961 / 5744421 / 0.946
59	3.3 M	mixed	freq.words	p5:1	10537 / 16721 / 0.630	5422300 / 5746045 / 0.944
60	0.8 M	mixed	words	p5:4	31160 / 82942 / 0.376	1469442 / 1645927 / 0.893
61	0.8 M	mixed	words	p5:3	29502 / 81154 / 0.364	1464979 / 1645072 / 0.89
62	0.8 M	mixed	words	p5:2	27573 / 78006 / 0.353	1454791 / 1639727 / 0.887
63	0.8 M	mixed	words	p5:1	27244 / 79884 / 0.341	1445265 / 1642051 / 0.88

Figure 1


sub-
corpus 1
(time t)
sub-
corpus 2
(time t′)
list of
FL-items
proportion
of shared
FL
list of
FL-items

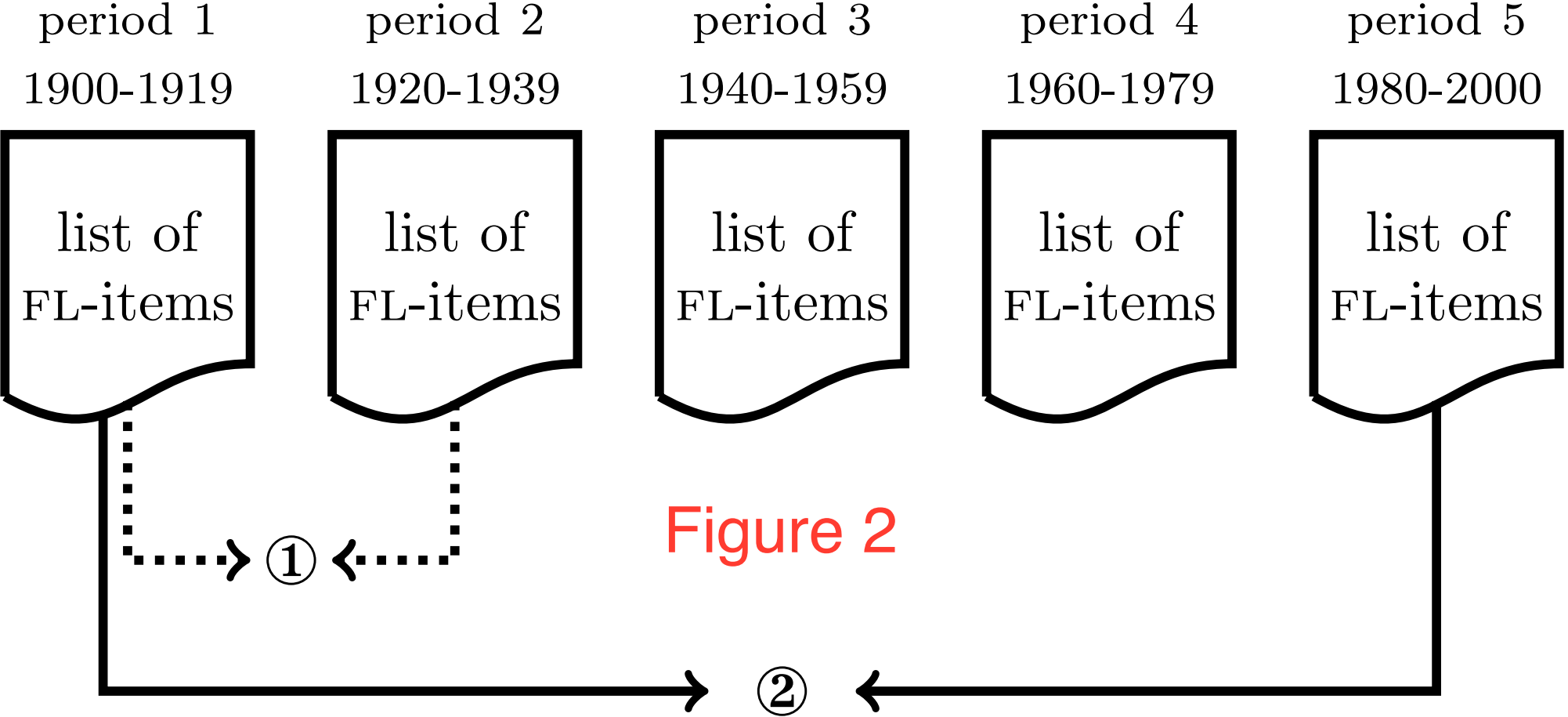

period 1
1900-1919
period 2
1920-1939
period 3
1940-1959
period 4
1960-1979
period 5
1980-2000
list of FL-items
list of FL-items
list of FL-items
list of FL-items
list of FL-items
①
②

Figure 2

Figure 3

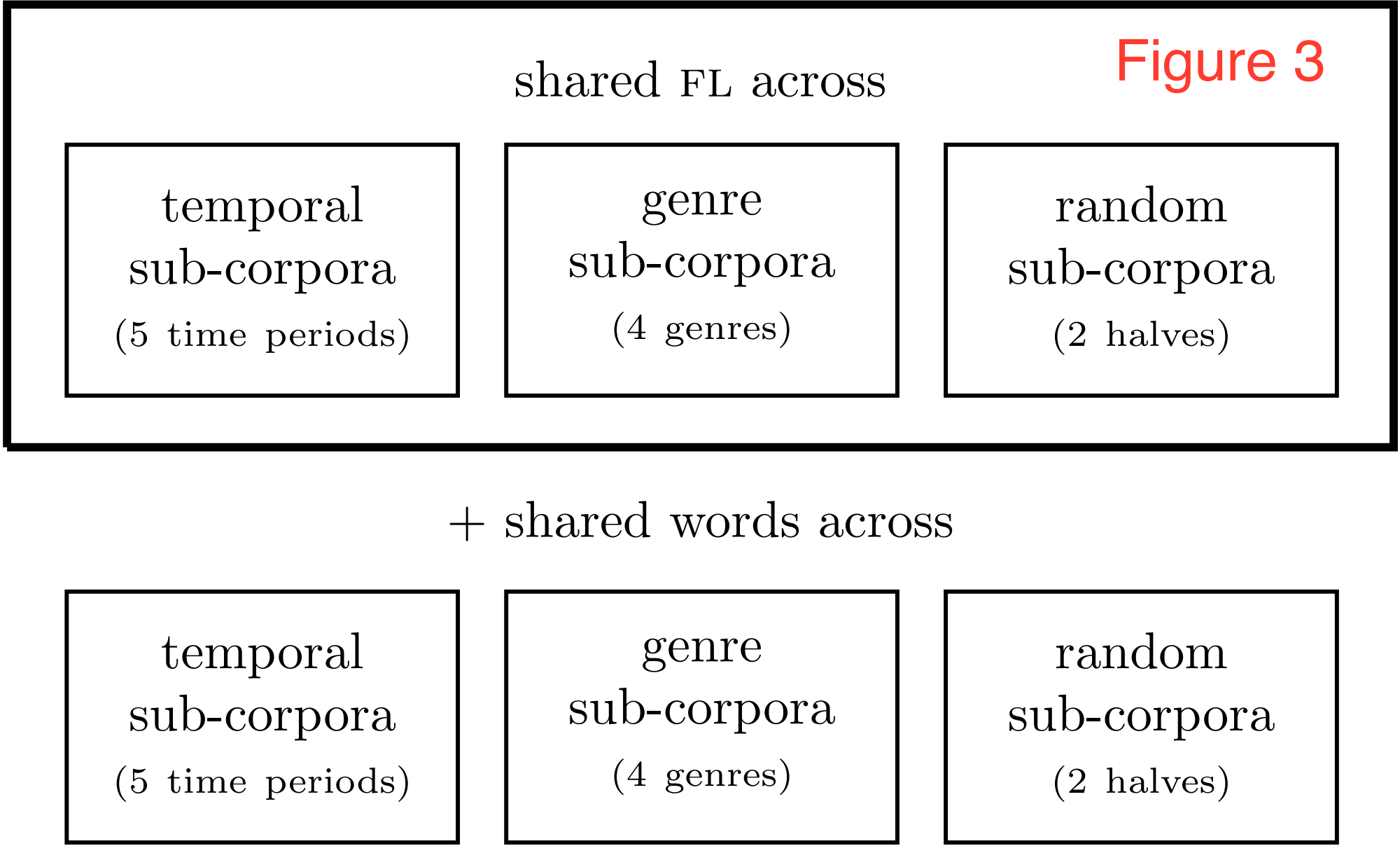

shared FL across
temporal sub-corpora (5 time periods)
genre sub-corpora (4 genres)
random sub-corpora (2 halves)
+ shared words across
temporal sub-corpora (5 time periods)
genre sub-corpora (4 genres)
random sub-corpora (2 halves)

Figure 4

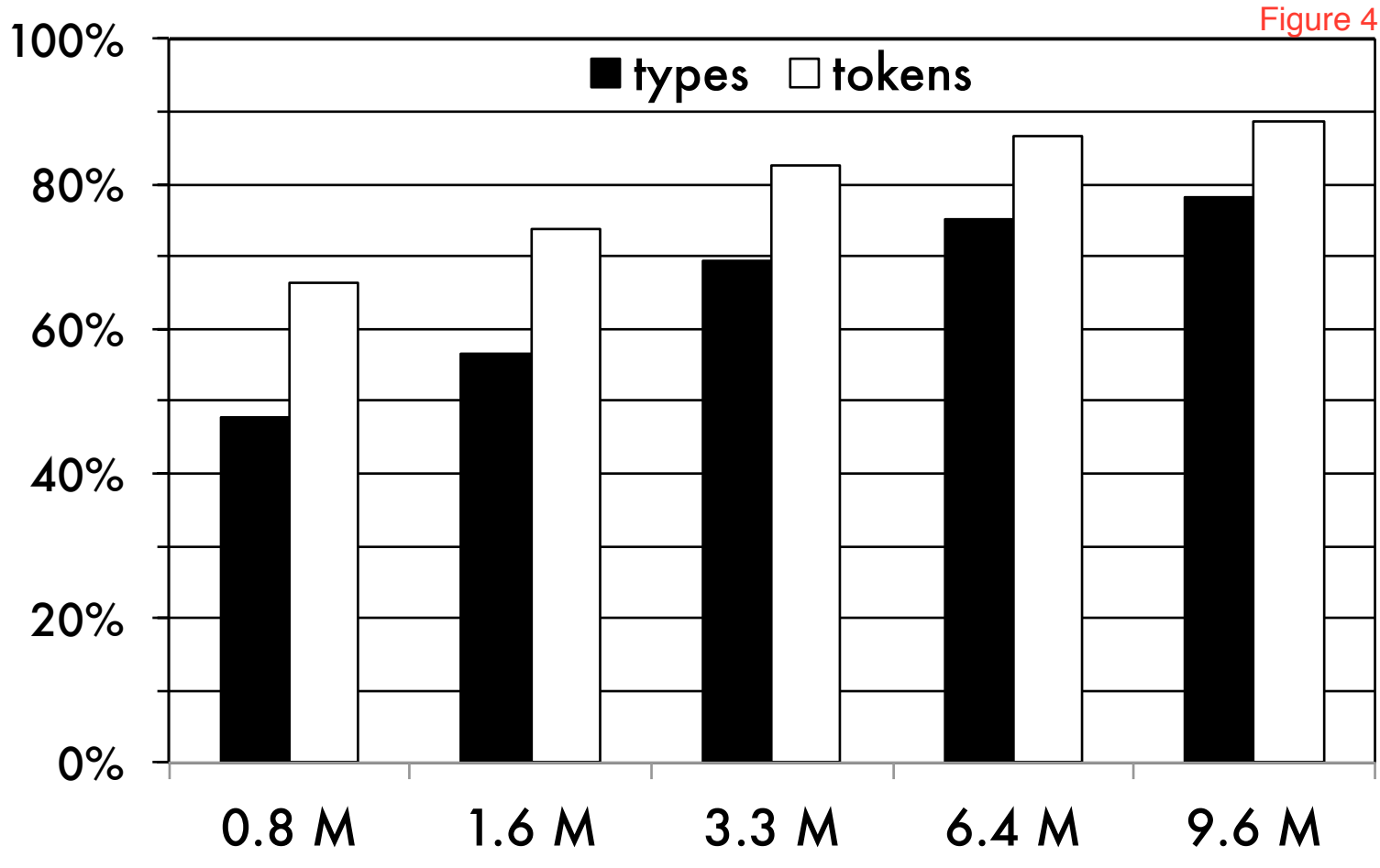


Figure 5

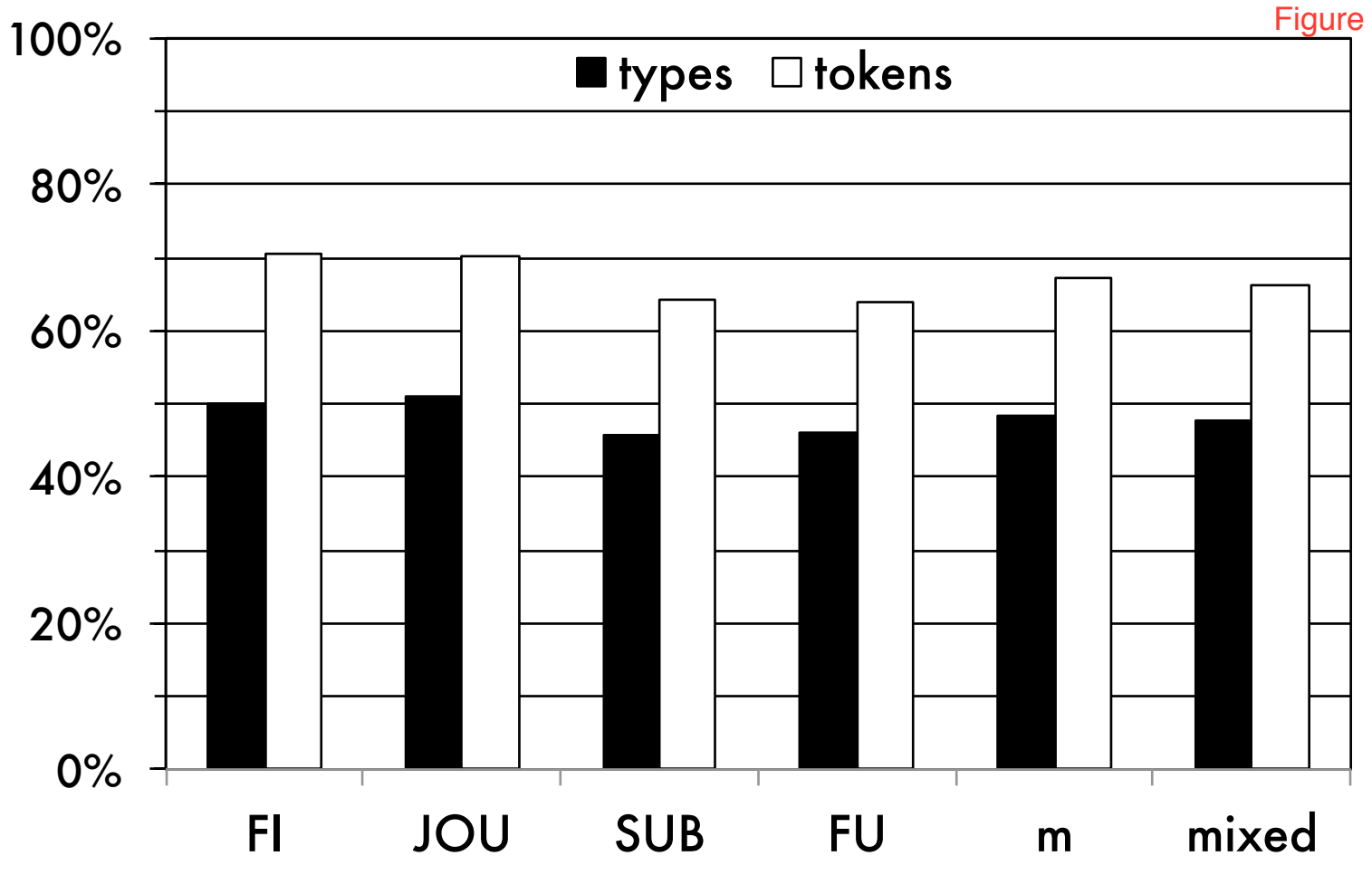


Figure 6

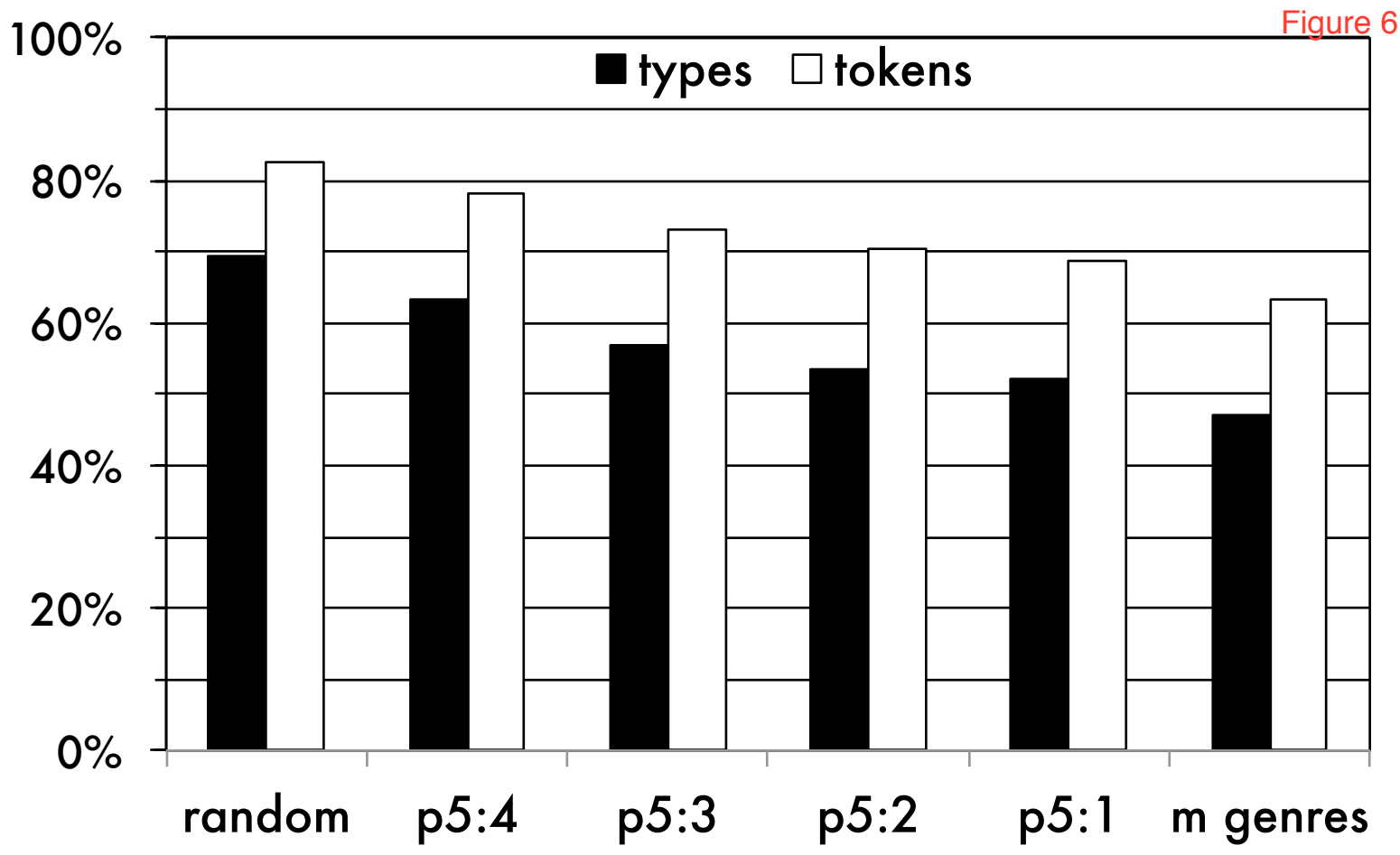


Figure 7

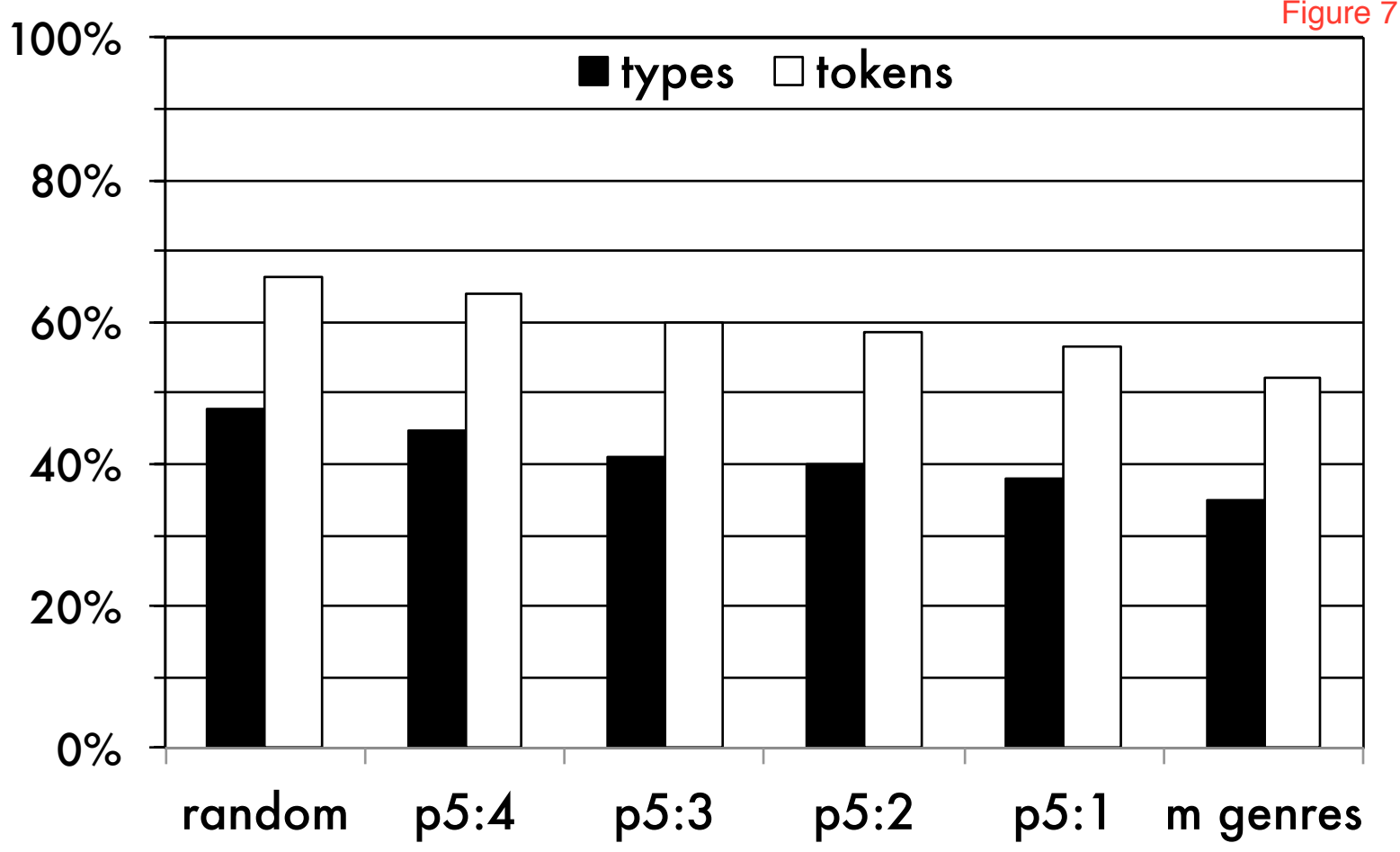


Figure 8

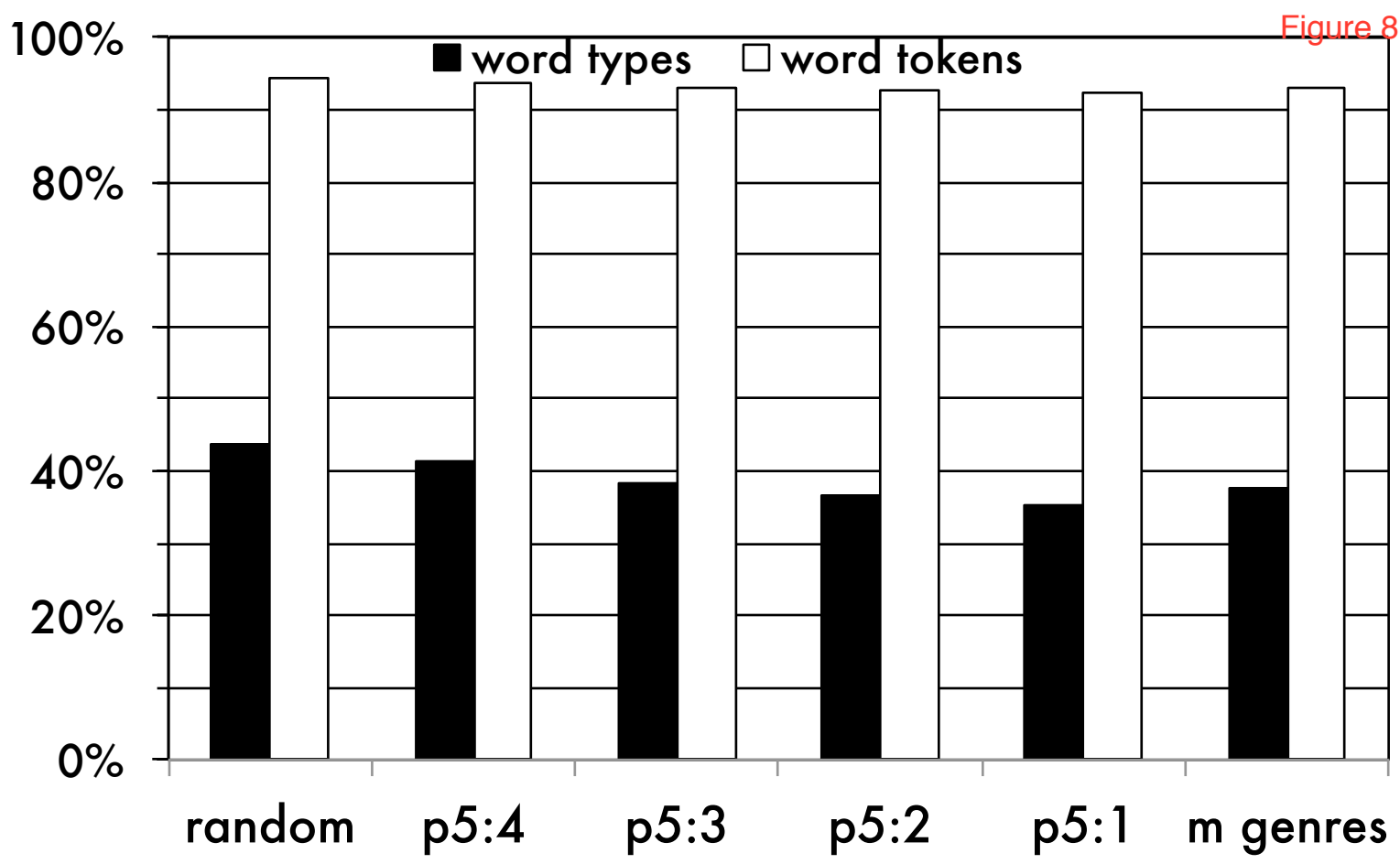

100%
80%
60%
40%
20%
0%
word types
word tokens
random
p5:4
p5:3
p5:2
p5:1
m genres

Figure 9

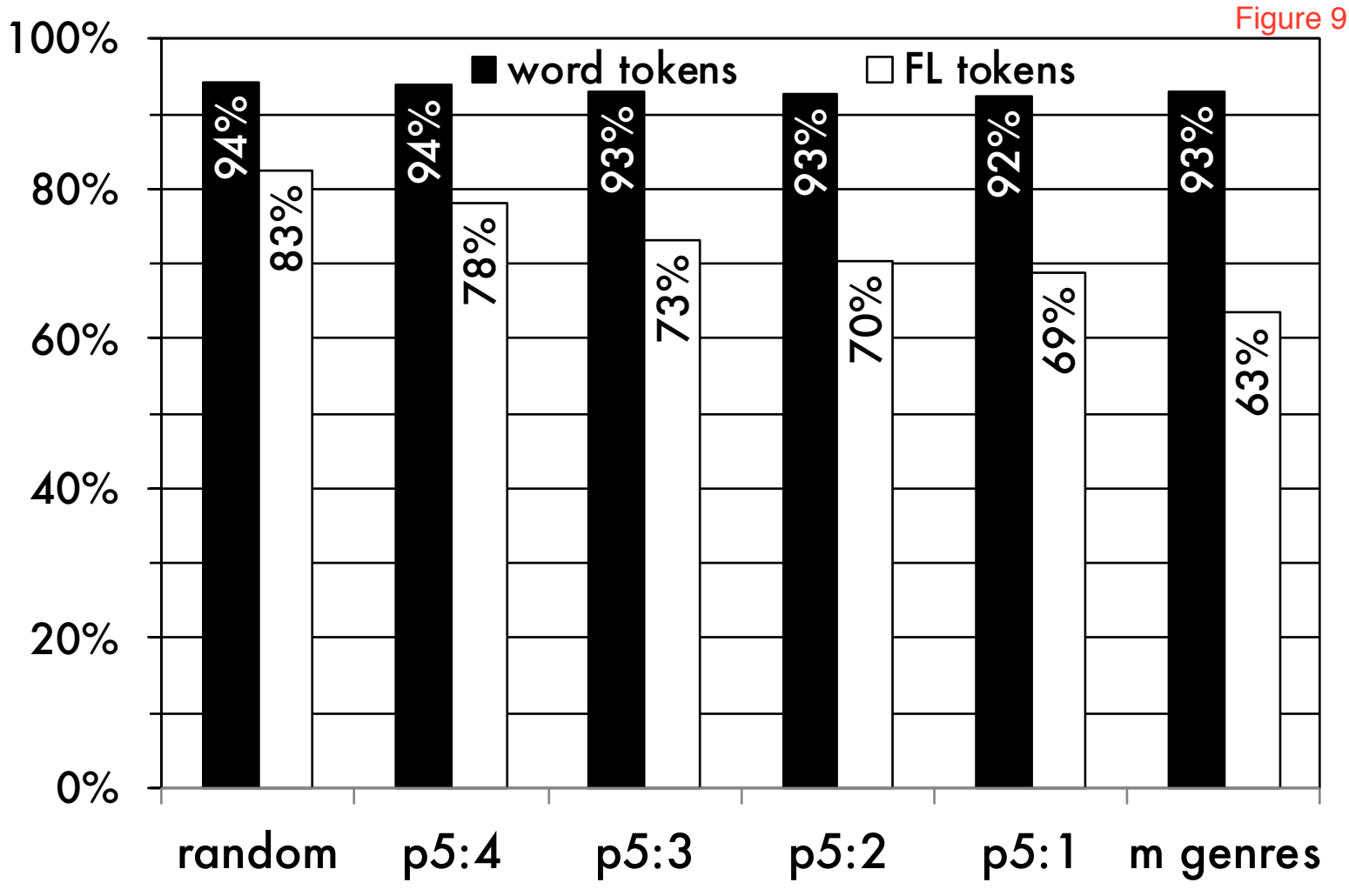
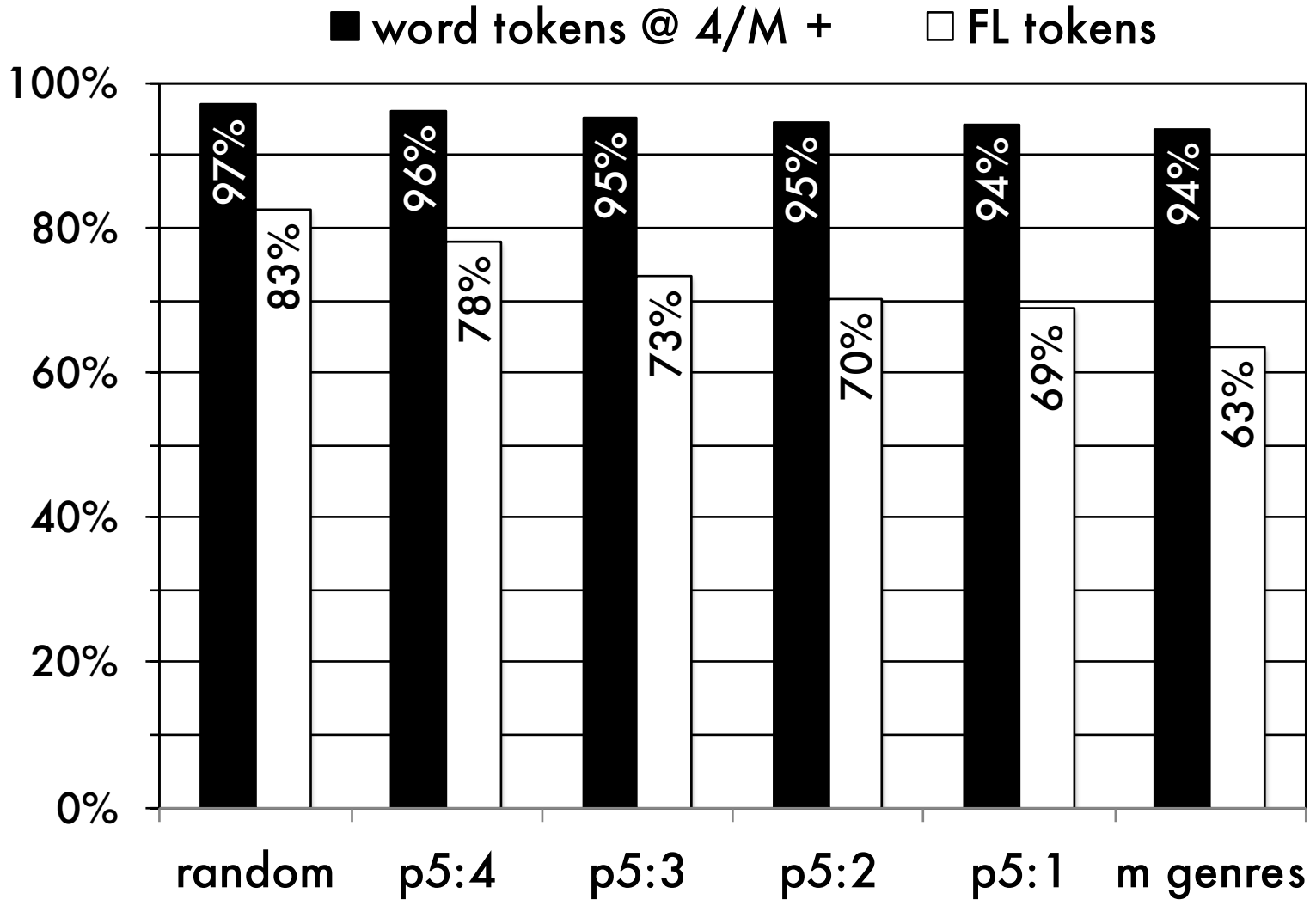

100%
80%
60%
40%
20%
0%
word tokens
FL tokens
94%
83%
94%
78%
93%
73%
93%
70%
92%
69%
93%
63%
random
p5:4
p5:3
p5:2
p5:1
m genres

Figure 10


word tokens @ 4/M +
FL tokens
100%
80%
60%
40%
20%
0%
97%
83%
96%
78%
95%
73%
95%
70%
94%
69%
94%
63%
random
p5:4
p5:3
p5:2
p5:1
m genres